Appendix A

Technical Details about the Generalized Exponential Model (GEM)

Appendix A

Technical Details about the Generalized Exponential Model (GEM)

A.1 Distance Function

Let $\Delta(w,d)$ denote the distance between the initial weights $d=\{d_k : k\in s\}$ and the adjusted weights *w*, with *k* being the k^{th} unit in the sample, and *s*, the sample selected. The distance function minimized under the generalized exponential model (GEM), subject to calibration constraints, is given by

$$\Delta(w,d)=\sum_{k\in s}\frac{d_k}{A_k}\left\{(a_k-\ell_k)\log\frac{a_k-\ell_k}{c_k-\ell_k}+(u_k-a_k)\log\frac{u_k-a_k}{u_k-c_k}\right\} \tag{A1.1}$$

where $a_k=w_k/d_k, A_k=(u_k-\ell_k)/[(u_k-c_k)(c_k-\ell_k)]$ and ℓ_k, c_k, and u_k are prescribed real numbers. Let T_x denote the *p*-vector of control totals corresponding to predictor variables $(x_1, ..., x_p)$. Then, the calibration constraints for the above minimization problem are

$$\sum_{k\in s} x_k d_k a_k = T_x \tag{A1.2}$$

The solution of the above minimization problem, if it exists, is given by a GEM with model parameters λ, i.e.,

$$a_k(\lambda)=\frac{\ell_k(u_k-c_k)+u_k(c_k-\ell_k)\exp\{A_k x_k'\lambda\}}{(u_k-c_k)+(c_k-\ell_k)\exp\{A_k x_k'\lambda\}} \tag{A1.3}$$

Note that the number of parameters in GEM should be $\le n$, where *n* is the size of the sample *s*. This is also the dimension of vectors *d* and *w*. It follows from Equation A1.3 that

$$\ell_k < a_k < u_k, k=1,\ldots,n \tag{A1.4}$$

The usual raking-ratio method (see, e.g., Singh & Mohl, 1996) of weight adjustment is a special case of GEM, such that for $\ell_k=0, u_k=\infty, c_k=1,$ and $k=1,\ldots,n$, we have

$$\Delta(w,d)=\sum_{k\in s} d_k a_k \log a_k - \sum_{k\in s} d_k (a_k-1) \tag{A1.5}$$

and $a_k(\lambda)=\exp(x_k'\lambda)$.

The logit method of Deville and Särndal (1992) is also a special case of GEM by setting $\ell_k = \ell, u_k = u,$ and $c_k = 1$ for all *k*.

A.2 GEM Adjustments for Extreme-Value Treatment, Nonresponse, and Poststratification

By choosing the user-specified parameters ℓ_k, c_k, and u_k appropriately, the unified GEM formula (A1.3) can be justified for all three types of adjustment. Denote the winsorized weights by $\{b_k\}$ where $b_k = d_k$ if d_k is not an extreme weight, and $b_k = med\{d_k\} \pm 3*$ IQR (where IQR denotes the interquartile range) if d_k is an extreme weight (where the quartiles for the weights are defined with respect to a suitable design-based stratum).

For the nonresponse adjustment, the sample is first divided into two parts: s^*, the non-extreme weight subsample; and s^{**}, the extreme weight subsample. For non-extreme weights, the following are set: $\ell_2 = 1, c_2 = \rho^{-1}, u_2 = u > \rho^{-1}$, where ρ is the overall response propensity; and for extreme weights with high weights, they are $\ell_k = \ell m_k, c_k = \rho^{-1} m_k, u_k = u_1 m_k$, where $m_k = b_k / d_k$ and $1 \le \ell_1 < \rho^{-1} = c_1 < u_1$ are prescribed numbers. Similarly, for extreme weights with low weights, $\ell_k = \ell_3 m_k, c_k = \rho^{-1} m_k, u_k = u_3 m_k$, and $1 \le \ell_3 < \rho^{-1} = c_3 < u_3$.

For the poststratification adjustment, for non-extreme weights, $\ell_k = \ell_2$, $c_k = c_2 = 1, u_k = u_2$, and for high extreme weights, $\ell_2 = \ell_1 m_k, c_k = m_k, u_k = u_1 m_k$, and, similarly for low extreme weights, $\ell_k = \ell_3 m_k, c_k = m_k, u_k = u_3 m_k$. The extreme-value adjustment is identical to poststratifcation, except for tighter bounds on extreme weights resulting from the final poststratification.

Notice that GEM allows the flexibility of specifying different bounds for different subsamples; in addition, the lower bound (in the case of nonresponse adjustments) can be made to equal one by choosing the center $c_k > 1$.

A.3 Newton-Raphson Steps

Let *X* denote the *n* x *p* matrix of predictor values, and for the v^{th} iteration,

$$\Gamma_{\phi v} = \text{diag}\left(d_k \phi_k^{(v)}\right), \phi_k^{(o)} = 1,$$

where

$$\phi_k^{(v)} = \left[\left(u_k - a_k^{(v)}\right)\left(a_k^{(v)} - \ell_k\right)\right] \Big/ \left[\left(u_k - c_k\right)\left(c_k - \ell_k\right)\right]$$

then, for Newton-Raphson iteration v, the value of the p-vector λ is adjusted as

$$\gamma^{(v)} = \gamma^{(v-1)} + \left(X' \Gamma_{\phi,v-1} X \right)^{-1} \left(T_x - \hat{T}_x^{(v-1)} \right)$$

where $\lambda^{(0)} = 1$.

The convergence criterion is based on the Euclidean distance $\left\| T_x - \hat{T}_x^{(v)} \right\|$. At each iteration, it is checked to determine whether it is decreasing or not. If not, a half-step is used in the iteration increment.

A.4 Scaled Constrained Exponential Model

In previous National Household Surveys on Drug Abuse (NHSDAs), constrained exponential models were used for poststratification and scaled constrained exponential models for nonresponse adjustments. The term "constrained exponential model" refers to the logit model of Deville and Särndal (1992) in which lower and upper bounds do not vary with k (i.e., $\ell_k = \ell, u_k = u,$ and $c_k = c = 1$ such that $\ell < 1 < u$. Thus, it is a special case of GEM. For the nonresponse adjustment, Folsom and Witt (1994) modified the constrained exponential models' estimating equations by a scaling factor (ρ^{-1}, the inverse of the overall response propensity) such that $1 < \rho^{-1} a_k < \rho^{-1} u$. This implies that choosing ℓ in constrained exponential models as ρ ensures that the scaled adjustment factor for nonresponse is at least one.

Appendix B

Derivation of Poststratification Control Totals

Appendix B
Derivation of Poststratification Control Totals

Unlike the person-level poststratification adjustment, the control totals for questionnaire dwelling unit (QDU)–level and person pair–level weight calibration cannot be derived from the U.S. Census directly. Estimates of the number of households and person pairs are not available at the domains we would like to control, and person-pair population estimates are not available even at a national level. However, by taking advantage of the two-phase design of the National Survey on Drug Use and Health (NSDUH), the screener dwelling unit (SDU) sample weights can be poststratified to U.S. Census population estimates. The calibrated SDU weights then can be used as stable control totals for the QDU- and person pair–level sample weights. In addition to the SDU weights, the person pair–level weights are calibrated to a second set of controls derived from the questionnaire, called household-level person counts. These controls are applied to pairs that are members of the 10 selected pair domains given below.

1. Parent-child pairs, child aged 12-14, target population is parents whose children aged 12-14 live with them;
2. Parent-child pairs, child aged 12-14, target population is children aged 12-14 living with their parents;
3. Parent-child pairs, child aged 12-17, target population is parents whose children aged 12-17 live with them;
4. Parent-child pairs, child aged 12-17, target population is children aged 12-17 living with their parents;
5. Parent-child pairs, child aged 12-20, target population is parents whose children aged 12-20 live with them;
6. Parent-child pairs, child aged 12-20, target population is children aged 12-20 living with their parents;
7. Sibling-sibling pairs, older sibling aged 15-17, younger sibling aged 12-14, target population is siblings aged 15-17 whose siblings are aged 12-14;
8. Sibling-sibling pairs, older sibling aged 18-25, younger sibling aged 12-17, target population is siblings aged 18-25 whose siblings are aged 12-17;
9. Spouse-spouse and partner-partner pairs; and
10. Spouse-spouse and partner-partner pairs with children under the age of 18 living in the household.

B.1 Derivation of QDU–Level Poststratification Controls

The derivation of QDU-level poststratification controls is not directly possible. Instead it must be based on work done for the person-level calibration. At the person level, weights are

calibrated to the control totals we wish to reach. These weights are then altered in order to conform to use with QDU-level data.

B.1.1 Person Level

B.1.1.1 Receiving and Deriving Person-Level Poststratification Control Totals

Civilian, noninstitutionalized population estimates for ages 12 and older are provided by the Population Estimates Branch of the U.S. Bureau of the Census. We receive two files, one at the national level and the other at the State level, each containing estimates of the population broken down by levels of month (12 levels), Hispanicity (2), race (6), sex (2), and age (6).

The breakdown received from the Census does not match the levels of the domains we would like to control. To account for this, we collapse levels. From this altered data, we create data sets with model group specific control totals. Observations in these data sets correspond to a breakdown by quarter (4), Hispanicity (2), race (5), sex (2), age (6), and number of States[15] in the model group (number of States varies according to which Census Region is represented in the model group).

B.1.1.2 Adjusting Screener Dwelling Unit Data to the Control Totals

In the person-level weighting, the SDU weights are poststratified to meet control totals based on the population estimates received from the Census. For NSDUH weighting, GEM is utilized to calibrate sample weights to multiple control totals. In doing so, each SDU receives an adjustment factor which, when multiplied by the initial weight, produces a final weight. The sum of all final weights corresponds to the civilian, noninstitutionalized population estimate for ages 12 and older, and the sum of all final weights in a domain corresponds to the control total for that domain. Note that there are a number of controls being calibrated to for each SDU, depending upon the domains to which the SDU belongs. The adjusted SDU weight reflects the civilian, noninstitutionalized population estimates for ages 12 and older, and can be utilized as a basis for constructing controls at the QDU and person pair levels.

B.1.2 QDU Level

B.1.2.1 Deriving QDU-Level Poststratification Control Totals from Adjusted SDU Weights

Since there are no controls for QDU-level poststratification available directly, we use the adjusted SDU weights. For these weights to be applicable at the QDU level, the SDU-level data must be restructured by sorting and summing over the domains to be used in the QDU-level calibration. This provides a data set where the summed weight, which still adds up to the proper population, is available for every domain to be utilized in the QDU calibration and, thus, can be used as a control total.

[15] The District of Columbia is included among States.

B.1.2.2 Adjusting QDU–Level Data to the Control Totals

As was done for the SDU data, the QDU-level data is adjusted via calibration in GEM of sample weights to multiple control totals. Each QDU receives an adjustment factor, similar to that described for the SDU weight in B1.1.2. The controls utilized in this calibration are based on the SDU weight as described in B.1.2.1 above. The adjusted weight is representative of the civilian, noninstitutionalized population estimates for ages 12 and older for all domains controlled within the modeling.

B.2 Derivation of Person Pair–Level Poststratification Controls

B.2.1 Deriving Person Pair–Level Poststratification Control Totals from Adjusted SDU Weights and the Household-Level Person Counts

Analogous to the QDU weights, some of the person pair controls are based on the SDU weights. However, two sets of control totals were utilized in the modeling, with one set based on the SDU weights and the other set based on the questionnaire roster.

For most pair data domains, those other than the 10 pair domains based on relationship, the control totals for the poststratification adjustments were obtained from SDU data, and were based on the number of possible pairs within SDUs. In order to obtain these pair counts belonging to various sociodemographic domains, the screener roster information was used to calculate all possible pairs within SDUs. For example, consider an SDU with two persons aged 12 to 17, and three persons aged 26 to 34. From this household composition, one can construct a single pair of persons aged 12 to 17, three pairs of persons aged 26 to 34, and six pairs of persons aged 12 to 17 and 26 to 34. It follows that the total number of possible pairs in this SDU is 10, from which the number of pairs belonging to the domain of interest can be obtained.

On the other hand, for the 10 selected pair domains based on relationship, the control totals for the poststratification adjustments were obtained from the questionnaire roster. This involved calibrating the pair weights to the number of persons in households belonging to each domain of interest. These controls were obtained from the larger sample of singles and pairs (i.e., one or two persons selected from DUs), and were calculated at the QDU (household) level. The pair weights were adjusted by the appropriate multiplicity. See Section 6.3 for details on the multiplicity counts and Section 6.4 on the household-level control totals, which are referred to there as household-level person counts.

B.2.2 Adjusting Person–Pair Level Data to the Control Totals

Like the SDU- and QDU-levels, the person pair-level data is adjusted via GEM. The use of two different types of controls requires a minor modification to the GEM macro so that both sets of controls may be addressed simultaneously. Similar to the SDU- and QDU-level poststratification steps, each pair receives an adjustment factor which, when multiplied by the initial weight, produces a final weight. The sum of all final weights corresponds to the civilian, noninstitutionalized population estimate for ages 12 and older, and the sum of all final weights in a domain corresponds to the control total for that domain.

Appendix C

GEM Modeling Summary for the Questionnaire Dwelling Unit Weights

Appendix C

GEM Modeling Summary for the Questionnaire Dwelling Unit Weights

Introduction

This appendix summarizes each questionnaire dwelling unit (QDU) model group throughout all stages of weight calibration modeling. Unlike much of the other information presented in this report, this section provides a model-specific overview of weight calibration, as opposed to a State- or domain-specific one.

For 2002, modeling involved taking four model groups through three adjustment steps: (1) selected dwelling unit poststratification, (2) respondent dwelling unit nonresponse adjustment, and (3) respondent dwelling unit poststratification. After the final poststratification, the adjusted sampling weights were reasonably distributed and did not require the additional treatment of the ev step.

Model-specific summary statistics are shown in Tables C1a, C1b to C4a, and C4b. Included in these tables, for each stage of modeling, are: the number of factor effects included; the high, low, and nonextreme weight bounds set to provide the upper and lower limits for the generalized exponential model (GEM) macro; weighted, unweighted and winsorized weight proportions; the unequal weighting effect (UWE); and weight distributions. The UWE provides an approximate partial measure of variance and provides a summary of how much impact a particular stage of modeling has on the distribution of the new product of weights. For more details on bounds, see Section 4.1. At each stage in the modeling, these summary statistics were calculated and utilized to help evaluate the quality of the current weight component under the model chosen.

Occurrences of small sample sizes and exact linear combinations in the realized data lead to situations whereby inclusion of all originally proposed levels of covariates in the model is not possible. The text and exhibits in Sections C1 to C4 summarize the decisions made with regard to final covariates included in each model. For a list of the proposed initial covariates considered at each stage of modeling, see Exhibit C.1; for the list of realized final model covariates, see Exhibits C1.1 to C4.3. The following sections establish a series of guidelines to assist in their interpretation.

C.1 Final Model Explanatory Variables

For brevity, numeric abbreviations for factor levels are established in Exhibit 4.1 (included here as Exhibit C.1 for easy reference) in Chapter 4. There, a complete list is provided of all variables and associated levels used at any stage of modeling. Note that not all factors or levels are present in all stages of modeling, and the initial set of variables is the same across model groups but changes for each stage of modeling. The initial candidates are found in any of

the proposed variables columns for a particular stage of weight adjustment. Exhibits C1.1 to C4.3 provide lists of the proposed and realized covariates.

To help understand what effects are controlled for at each stage of the modeling, it may be useful to create cross-classification tables as shown in Section C.3. Sections C.2 and C.3 explain how to use various exhibits for selected model variables to construct these tables.

Exhibit C.1 Definitions of Levels for QDU-Level Calibration Modeling Variables

Age [c]
1: 12-17, 2: 18-25, 3: 26-34, 4: 35-49, 5: 50+ [1]

Gender [c]
1: Male, 2: Female [1]

Group Quarter Indicator [b]
1: College Dorm, 2: Other Group Quarter, 3: Nongroup Quarter [1]

Hispanicity [c]
1: Hispanic, 2: Non-Hispanic [1]

Household Size [c]
Continuous variable - count of individuals rostered with DU.

Household Type (ages of persons rostered within DU) [b]
1: 12-17, 18-25, 26+, 2: 12-17, 18-25, 3: 12-17, 26+, 4: 18-25, 26+, 5: 12-17, 6: 18-25, 7: 26+

Percentage of Owner-Occupied Dwelling Units in Segment (% Owner) [b]
1: 50%-100% [1], 2: 10%->50%, 3: 0->10%

Percentage of Segments That Are Black (% black) [b]
1: 50%-100%, 2: 10%->50%, 3: 0->10% [1]

Percentage of Segments That Are Hispanic (% Hispanic) [b]
1: 50%-100%, 2: 10%->50%, 3: 0->10% [1]

Population Density [b]
1: MSA 1,000,000 or more, 2: MSA less than 1,000,000, 3: Non-MSA urban, 4: Non-MSA rural [1]

Quarter [bc]
1: Quarter 1, 2: Quarter 2, 3: Quarter 3, 4: Quarter 4 [1]

Race (3 levels) [c]
1: white, [1] 2: black, 3: Other

Race (4 levels) [c]
1: white, [1] 2: black, 3: Native American, 4: Asian, 5: multi-race.

Race of Householder [b]
1: Hispanic white [1], 2: Hispanic black, 3: Hispanic others, 4: Non-Hispanic white,
5: Non-Hispanic black, 6: Non-Hispanic others,

Relation to Householder [b]
1: Householder or Spouse, 2: Child, 3: Other Relative, 4: Nonrelative [1]

Segment Combined Median Rent and Housing Value (Rent/Housing) [b2]
1: First Quintile, 2: Second Quintile, 3: Third Quintile, 4: Fourth Quintile, 5: Fifth Quintile [1]

Exhibit C.1 Definitions of Levels for QDU-Level Calibration Modeling Variables (continued)

State [bc3]

Model Group 1: 1: Connecticut, 2: Maine, 3: Massachusetts, [1] 4: New Hampshire, 5: New Jersey, 6: New York, 7: Pennsylvania, 8: Rhode Island, 9: Vermont

Model Group 2: 1: Illinois, 2: Indiana, 3: Iowa, 4: Kansas, 5: Michigan, 6: Minnesota, 7: Missouri, 8: Nebraska, 9: North Dakota, 10: Ohio, 11: South Dakota, 12: Wisconsin [1]

Model Group 3: 1: Alabama, 2: Arkansas, 3: Delaware, 4: District of Columbia, 5: Florida, 6: Georgia, 7: Kentucky, 8: Louisiana, 9: Maryland, 10: Mississippi, 11: North Carolina, [1] 12: Oklahoma, 13: South Carolina, 14: Tennessee, 15: Texas, 16: Virginia, 17: West Virginia

Model Group 4: 1: Alaska, 2: Arizona, [1] 3: California, 4: Colorado, 5: Idaho, 6: Hawaii, 7: Montana, 8: Nevada, 9: New Mexico, 10: Oregon, 11: Utah, 12: Washington, 13: Wyoming

State/Region [b3]

Model Group 1: 1: New York, 2: Pennsylvania, 3: other [1]

Model Group 2: 1: Illinois, 2: Michigan, 3: Ohio, 4: other [1]

Model Group 3: 1: Florida, 2: Texas, 3: other [1]

Model Group 4: 1: California, 2: other [1]

[1] The reference level for this variable. This is the level against which effects of other factor levels are measured.

[2] Segment-combined Median Rent and Housing Value is a composite measure based on rent, housing value, and percent owner occupied.

[3] The States or district assigned to a particular model is based on Census regions.

[b] Binary variable.

[c] Counting variable. A count of all persons in the household.

C.2 Glossary of Terms Used in the Description of the Variables in the Final Model

Factor effect. Represents the effects of levels considered for one-factor, two-factor, and higher order factors.

Reference/reference set. Factor effects composed of reference levels are not explicitly listed in the set of model variables. However, these effects manifest themselves either separately or in combination with other factors depending on the presence of other factors in the model.

All levels present. All effects and all levels of the factor under consideration are in the model.

Coll. *(levels)*. Collapse these factor effects together. Factor effects that have been collapsed with others manifest themselves jointly in the model.

Drop all levels. All factor effects are completely removed from the model for all levels and any combinations involving this factor.

Drop *level(s)*. Collapse these factor effects into the reference set. The factor effects comprising the dropped levels are manifested jointly with either some of or all of the factor effects in the reference set.

Drop *level(s)*; sing. During the modeling process the factor effects listed were removed from the model due to singularity.

Drop *level(s); zero cnts.* During the modeling process the factor effects listed were removed from the model due to zero sample.

Hier. One or more of the effects in a higher order interaction was collapsed or dropped in an interaction at a lower order, either eliminating or combining factors of higher order interactions with that effect.

Do the same for *(effects)*. Repeat the previous step for all effect levels listed.

Drop or Collapse using*. The asterisk is used as a wildcard character to indicate all levels of the factor for that effect.

*Note: The above are given as a list of general terms. Certain other specific terms are sometimes used within a particular section.

C.3 How to Interpret Collapsing and Dropping of Factor Effects

To help visualize what effects are directly controlled for in our model, one can construct the table that reflects the collapsing scheme employed. The following is a complex example from the 1999 person-level modeling.

1. Locate the Factor effect - Model 9 Person Nonresponse Adjustment:

Three-Factor Effects	**Comments**
State × Age × Race (3 Levels)	Coll. (2,1,2) & (2,1,3); hier. Repeat for all levels of age in State (2); hier. Drop (3,4,2); sing. Collapse (1,4,2) & (1,4,3); conv. Drop (3,*,*); conv. Coll. (4,1,2) & (4,1,3); conv. Repeat for all levels of age in State (4).

2. Determine the initial range of possible levels for the variables by referring to the variable definitions. See Exhibits C.1 and H.1 for QDU- and Pair-level variable definitions. In addition, the columns 'Levels,' 'Proposed,' and 'Final' will provide counts of all factor effects, all explicitly proposed factors, and all explicitly controlled factors, but these are not necessary for construction of the cross-classification table. The following example is based upon person-level variables, but the process is the same.

- State (for the model group in question, in this case, Model Group 9)
Model Group 9: 1: Alaska, 2: Hawaii, 3: Oregon, 4: Washington, 5: California[1,2,3]

- Age
1: 12 to 17,[2,3] 2: 18 to 25, 3: 26 to 34, 4: 35 to 49, 5: 50+ [1]

- Race (3 levels)
1: white,[1,2,3] 2: black, 3: other

Note that superscript number indicates the reference level of the variable for a particular stage of modeling. In our case, the model stage is 'Person Nonresponse Adjustment.'

3. Construct the cross-classification table.

For example, Race (4 Levels) is defined this way:

Race (4 Levels	white	black	Asian	American Indian/Alaska Native

Indicates the reference-level set.

This is the cross-classification table for State × Race (4 Levels):

State*Race (4 Levels)	white	black	Asian	American Indian/Alaska Native
AK				
HI				
OR				
WA				
CA				

Indicates the reference-level set.

The cross-classification table of interest (State × Age × Race [3 Levels]) is as follows:

State × Age × Race (3 Levels)	white	black	other
AK × 12-17			
18-25			
26-34			
35-49			
50+			
HI × 12-17			
18-25			
26-34			
35-49			
50+			
OR × 12-17			
18-25			
26-34			
35-49			
50+			
WA × 12-17			
18-25			
26-34			
35-49			
50+			
CA × 12-17			
18-25			
26-34			
35-49			
50+			

Indicates the reference-level set.

The number of respondents in that class at this stage of modeling would appear within each cell of the table. Construction of the other cross-classification tables follows the same logic and is only necessary to the point of providing understanding of the final table.

4. Use the information under the 'Comments' column definition to determine the combination of factors controlled.

Hier. This note means the factor effect was collapsed at a lower order. Because this note is present, examine the information on lower-order factor effects that are the components of the interaction term, State × Race (3 levels) × Age; that is, look at the one-factor and two-factor effects for State, Race (4 levels) and Age, and their accompanying information:

One-Factor Effects	**Comments**
State	All levels present.
Race (4 Levels)	All levels present.
Age	All levels present.
Two-Factor Effects	**Comments**
State × Age	All levels present.
State × Race (4 Levels)	Collapse (1,3) & (1,4). Do the same for all other States except (2). Collapse (2,2), (2,3), & (2,4).
Age × Race (3 Levels)	All levels present.

The reason for the note is the State × Race (4 Levels) interaction. It indicates a need to maintain the collapsing scheme when setting up any three-factor crosses involving State × Race. Following these directions, the resulting two-factor table is:

State × Race (4 Levels)	white	black	Asian	American Indian/Alaska Native
AK				
HI				
OR				
WA				
CA				

Indicates the reference-level set.

Returning to our instructions, we see that several other factor crosses have been affected by modeling:

Three-Factor Effects	**Comments**
State × Age × Race (3 Levels)	Coll. (2,1,2) & (2,1,3); hier. Repeat for all levels of age in State (2); hier. Drop (3,4,2); sing. Collapse (1,4,2) & (1,4,3); conv. Drop (3,*,*); conv. Coll. (4,1,2) & (4,1,3); conv. Repeat for all levels of age in State (4).

Construct the complete table, then begin combining blocks as directed. The unshaded cells represent the factors directly controlled for by the model. The shaded cells represent the composite reference set, whose values may be obtained by utilizing the marginal sums, although when changes to the initially proposed set occur, it can make certain reference cell counts indistinguishable.

After following the directions, the cross-classification table should appear as follows:

State × Age × Race (3 Levels)	white	black	other
AK × 12-17			
18-25			
26-34			
35-49			
50+			
HI × 12-17			
18-25			
26-34			
35-49			
50+			
OR × 12-17			
18-25			
26-34			
35-49			
50+			
WA × 12-17			
18-25			
26-34			
35-49			
50+			
CA × 12-17			
18-25			
26-34			
35-49			
50+			

Indicates the reference-level set.

Exhibit C.2 Covariates for 2002 NSDUH Questionnaire Dwelling Unit Weights

Variables	Binary	Counting	Level	Proposed
One-Factor Effects		**76**	**76**	
Intercept	Y		1	1
Population density	Y		4	3
Group quarter	Y		3	2
Race of Householder	Y		6	5
Rent/housing value	Y		5	4
Segment % Black	Y		3	2
Segment % Hispanic	Y		3	2
Segment % Owner-Occupied	Y		3	2
Household type	Y		7	6
State	Y	Y	Model Specific	
Quarter	Y	Y	4	3
Age Group		Y	5	4
Race		Y	5	4
Hispanicity		Y	2	1
Gender		Y	2	1
Household size		Y	1	1
Two-Factor Effects				
Age x Race (3 levels)		Y	5 x 3	8
Age x Hispanicity		Y	5 x 2	4
Age x Gender		Y	5 x 2	4
Race (3 levels) x Hispanicity		Y	3 x 2	2
Race (3 levels) x Gender		Y	3 x 2	2
Hispanicity x Gender		Y	2 x 2	1
State x Age		Y	Model Specific	
State x Race (5 levels)		Y	Model Specific	
State x Gender		Y	Model Specific	
State x Hispanicity		Y	Model Specific	
% Black x % Owner	Y		3 x 3	4
% Black x Rent/housing		Y	3 x 5	8
% Hispanicity x % Owner		Y	3 x 3	4
% Hispanicity x Rent/housing		Y	3 x 5	8
% Owner x Rent/housing	Y		3 x 5	8
Three-Factor Effects				
Race (3 levels) x Age x Gender		Y	8	8
State/Region x Age x Gender		Y		
State/Region x Age x Hispanicity		Y		
State/Region x Age x Race (3 levels)		Y		
State/Region x Hispanicity x Gender		Y		
State/Region x Race (3 levels) x Hispanicity		Y		
State/Region x Race (3 levels) x Gender		Y		

Appendix C1

Model Group 1: Northeast

Connecticut, Maine, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, Vermont

Table C1a 2002 QDU Weight GEM Modeling Summary (Model Group 1: Northeast)

Modeling Step[1]	Extreme Weight Proportions			UWE[2]	# XVAR[3]	Bounds[4]	
	Unweighted	Weighted	Outwinsor			Nominal	Realized
sel.qdu.ps	2.01%	2.31%	0.50%	2.9194	243	(0.6, 3.2)	(0.60, 3.20)
	1.61%	2.28%	0.57%	2.9454	242	(0.6, 3.2)	(0.60, 3.20)
						(0.9, 3.2)	(0.90, 3.20)
res.qdu.nr	1.70%	2.45%	0.67%	2.9988	243	(1.0, 4.3)	(1.00, 4.30)
	1.49%	2.82%	0.75%	3.6350	241	(1.0, 4.3)	(1.00, 4.27)
						(1.1, 4.3)	(1.10, 1.68)
res.qdu.ps	1.49%	2.82%	0.75%	3.6350	243	(0.83, 2.10)	(0.87, 2.10)
	1.65%	2.41%	0.55%	3.6416	242	(0.84, 2.10)	(0.84, 1.97)
						(0.84, 1.14)	(0.84, 1.12)

[1] For a key to modeling abbreviations, see Chapter 7, Exhibit 7.1.

[2] Unequal weighting effect defined as $1+[(n-1)/n]*CV^2$, where CV=coefficient of variation of weights.

[3] Number of proposed covariates on top line, and number finalized after modeling.

[4] There are six sets of bounds for each modeling step. Nominal bounds are used in defining maximum/minimum values for the GEM adjustment factors. The realized bound is the actual adjustment produced by the modeling. The set of three bounds listed for each step correspond to the high extreme values, the nonextreme values, and the low extreme values.

Table C1b 2002 Distribution of Weight Adjustment Factors and Weight Products (Model Group 1: Northeast)

	SDU wt	QDU design weight		sel.qdu.ps		res.qdu.nr		res.qdu.ps	
	1-9	duwght10	1-10	duwght11	1-11	duwght12	1-12	duwght13	1-13
Minimum	20	1.00	20	0.34	14	0.24	14	0.47	12
1%	89	1.00	101	0.61	101	1.00	103	0.85	93
5%	145	1.00	167	0.77	163	1.00	174	0.92	172
10%	194	1.00	246	0.83	242	1.02	256	0.97	254
25%	360	1.00	557	0.91	540	1.06	568	0.99	564
Median	669	1.05	844	0.99	843	1.12	906	1.00	908
75%	947	3.34	1,767	1.09	1,785	1.23	1,870	1.01	1,891
90%	1,330	6.28	4,674	1.22	4,789	1.40	5,268	1.03	5,256
95%	1,582	9.03	7,127	1.34	7,190	1.55	9,086	1.04	8,999
99%	2,173	14.78	11,960	1.77	12,137	2.05	17,108	1.17	17,238
Maximum	8,298	19.05	27,935	13.64	38,525	4.29	41,057	1.97	40,953
n	11,436	-	11,436	-	11,436	-	9,724	-	9,724
Mean	732	2.50	1,776	1.02	1,807	1.18	2,126	1.00	2,126
Max/Mean	11.33	-	15.73	-	21.31	-	19.32	-	19

[1] For a key to modeling abbreviations, see Chapter 7, Exhibit 7.1.

Model Group 1 Overview

Selected Questionnaire Dwelling Unit–Level Poststratification

The Northeast model group maintained all originally proposed covariates except for one level of the segment percent Hispanic by segment percent owner-occupied variable, which was removed due to zero counts.

Respondent Questionnaire Dwelling Unit–Level Nonresponse

All main effects were maintained in full for the nonresponse adjustment, but among two-factor effects, the Massachusetts race variables Native American and Asian were collapsed. Zero sample also led to the removal of a level of the segment level variable, percent Hispanic by percent owner-occupied.

Respondent Questionnaire Dwelling Unit–Level Poststratification

As in the previous selected QDU-level poststratification and the respondent QDU-level nonresponse adjustment steps, the segment level variable percent Hispanic by percent owner-occupied had levels collapsed due to zero sample. All other proposed covariates were maintained in full.

Exhibit C1.1 Covariates for 2002 NSDUH Questionnaire Dwelling Unit Weights (sel.qdu.ps) Model Group 1: Northeast

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**60**	**60**	
Intercept	1	1	1	All levels present.
Group quarter	3	2	2	All levels present.
Race of Householder	6	5	5	All levels present.
Household Type	7	6	6	All levels present.
Household Size	1	1	1	All levels present.
Rent/Housing	5	4	4	All levels present.
Population Density	4	3	3	All levels present.
% Black	3	2	2	All levels present.
% Hispanic	35	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
State (count)	9	8	8	All levels present.
State (binary)	9	8	8	All levels present.
Quarter (count)	4	3	3	All levels present.
Quarter (binary)	4	3	3	All levels present.
Age Group	5	4	4	All levels present.
Race	5	4	4	All levels present.
Hispanicity	2	1	1	All levels present.
Gender	2	1	1	All levels present.
Two-Factor Effects		**133**	**132**	
Age x Race (3 levels)	5 x 3	8	8	All levels present.
Age x Hispanicity	5 x 2	4	4	All levels present.
Age x Gender	5 x 2	4	4	All levels present.
Race (3 levels) x Hispanicity	3 x 2	2	2	All levels present.
Race (3 levels) x Gender	3 x 2	2	2	All levels present.
Hispanicity x Gender	2 x 2	1	1	All levels present.
State x Age	9 x 5	32	32	All levels present.
State x Race	9 x 5	32	32	All levels present.
State x Gender	9 x 2	8	8	All levels present.
State x Hispanicity	9 x 2	8	8	All levels present.
% Black x % Owner-Occupied	3 x 3	4	4	All levels present.
% Black x Rent/Housing	3 x 5	8	8	All levels present.
% Hispanicity x % Owner-Occupied	3 x 3	4	3	Coll (1,1) & (2,1); zero cnts.
% Hispanicity x Rent/housing	3 x 5	8	8	All levels present.
% Owner-Occupied x Rent/housing	3 x 5	8	8	All levels present.
Three-Factor Effects		**50**	**50**	
Race (3 levels) x Age x Gender	3 x 5 x 2	8	8	All levels present.
State/Region x Age x Gender	3 x 5 x 2	8	8	All levels present.
State/Region x Age x Hispanicity	3 x 5 x 2	8	8	All levels present.
State/Region x Age x Race (3 levels)	3 x 5 x 3	16	16	All levels present.
State/Region x Hispanicity x Gender	3 x 2 x 2	2	2	All levels present.
State/Region x Race (3 levels) x Hispanicity	3 x 3 x 2	4	4	All levels present.
State/Region x Race (3 levels) x Gender	3 x 3 x 2	4	4	All levels present.
Total		**243**	**242**	

Exhibit C1.2 Covariates for 2002 NSDUH Questionnaire Dwelling Unit Weights (res.qdu.nr) Model Group 1: Northeast

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**60**	**60**	
Intercept	1	1	1	All levels present.
Group quarter	3	2	2	All levels present.
Race of Householder	6	5	4	All levels present.
Household Type	7	6	6	All levels present.
Household Size	1	1	1	All levels present.
Rent/Housing	5	4	4	All levels present.
Population Density	4	3	3	All levels present.
% Black	3	2	2	All levels present.
% Hispanic	35	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
State (count)	9	8	8	All levels present.
State (binary)	9	8	8	All levels present.
Quarter (count)	4	3	3	All levels present.
Quarter (binary)	4	3	3	All levels present.
Age Group	5	4	4	All levels present.
Race	5	4	4	All levels present.
Hispanicity	2	1	1	All levels present.
Gender	2	1	1	All levels present.
Two-Factor Effects		**133**	**131**	
Age x Race (3 levels)	5 x 3	8	8	All levels present.
Age x Hispanicity	5 x 2	4	4	All levels present.
Age x Gender	5 x 2	4	4	All levels present.
Race (3 levels) x Hispanicity	3 x 2	2	2	All levels present.
Race (3 levels) x Gender	3 x 2	2	2	All levels present.
Hispanicity x Gender	2 x 2	1	1	All levels present.
State x Age	9 x 5	32	32	All levels present.
State x Race	9 x 5	32	31	Coll (3,3) & (3,4); zero conv.
State x Gender	9 x 2	8	8	All levels present.
State x Hispanicity	9 x 2	8	8	All levels present.
% Black x % Owner-Occupied	3 x 3	4	4	All levels present.
% Black x Rent/Housing	3 x 5	8	8	All levels present.
% Hispanicity x % Owner-Occupied	3 x 3	4	3	Coll (1,1) & (2,1); zero cnts.
% Hispanicity x Rent/housing	3 x 5	8	8	All levels present.
% Owner-Occupied x Rent/housing	3 x 5	8	8	All levels present.
Three-Factor Effects		**50**	**50**	
Race (3 levels) x Age x Gender	3 x 5 x 2	8	8	All levels present.
State/Region x Age x Gender	3 x 5 x 2	8	8	All levels present.
State/Region x Age x Hispanicity	3 x 5 x 2	8	8	All levels present.
State/Region x Age x Race (3 levels)	3 x 5 x 3	16	16	All levels present.
State/Region x Hispanicity x Gender	3 x 2 x 2	2	2	All levels present.
State/Region x Race (3 levels) x Hispanicity	3 x 3 x 2	4	4	All levels present.
State/Region x Race (3 levels) x Gender	3 x 3 x 2	4	4	All levels present.
Total		**243**	**241**	

Exhibit C1.3 Covariates for 2002 NSDUH Questionnaire Dwelling Unit Weights (res.qdu.ps) Model Group 1: Northeast

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**60**	**60**	
Intercept	1	1	1	All levels present.
Group quarter	3	2	2	All levels present.
Race of Householder	6	5	5	All levels present.
Household Type	7	6	6	All levels present.
Household Size	1	1	1	All levels present.
Rent/Housing	5	4	4	All levels present.
Population Density	4	3	3	All levels present.
% Black	3	2	2	All levels present.
% Hispanic	35	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
State (count)	9	8	8	All levels present.
State (binary)	9	8	8	All levels present.
Quarter (count)	4	3	3	All levels present.
Quarter (binary)	4	3	3	All levels present.
Age Group	5	4	4	All levels present.
Race	5	4	4	All levels present.
Hispanicity	2	1	1	All levels present.
Gender	2	1	1	All levels present.
Two-Factor Effects		**133**	**132**	
Age x Race (3 levels)	5 x 3	8	8	All levels present.
Age x Hispanicity	5 x 2	4	4	All levels present.
Age x Gender	5 x 2	4	4	All levels present.
Race (3 levels) x Hispanicity	3 x 2	2	2	All levels present.
Race (3 levels) x Gender	3 x 2	2	2	All levels present.
Hispanicity x Gender	2 x 2	1	1	All levels present.
State x Age	9 x 5	32	32	All levels present.
State x Race	9 x 5	32	32	All levels present.
State x Gender	9 x 2	8	8	All levels present.
State x Hispanicity	9 x 2	8	8	All levels present.
% Black x % Owner-Occupied	3 x 3	4	4	All levels present.
% Black x Rent/Housing	3 x 5	8	8	All levels present.
% Hispanicity x % Owner-Occupied	3 x 3	4	3	Coll (1,1) & (2,1); zero cnts.
% Hispanicity x Rent/housing	3 x 5	8	8	All levels present.
% Owner-Occupied x Rent/housing	3 x 5	8	8	All levels present.
Three-Factor Effects		**50**	**50**	
Race (3 levels) x Age x Gender	3 x 5 x 2	8	8	All levels present.
State/Region x Age x Gender	3 x 5 x 2	8	8	All levels present.
State/Region x Age x Hispanicity	3 x 5 x 2	8	8	All levels present.
State/Region x Age x Race (3 levels)	3 x 5 x 3	16	16	All levels present.
State/Region x Hispanicity x Gender	3 x 2 x 2	2	2	All levels present.
State/Region x Race (3 levels) x Hispanicity	3 x 3 x 2	4	4	All levels present.
State/Region x Race (3 levels) x Gender	3 x 3 x 2	4	4	All levels present.
Total		**243**	**242**	

Appendix C2

Model Group 2: Midwest

Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, North Dakota, Ohio, South Dakota, Wisconsin

Table C2a 2002 QDU Weight GEM Modeling Summary (Model Group 2: Midwest)

Modeling Step[1]	Extreme Weight Proportions			UWE[2]	# XVAR[3]	Bounds[4]	
	Unweighted	Weighted	Outwinsor			Nominal	Realized
sel.qdu.ps	2.17%	1.61%	0.24%	2.7762	300	(0.3, 4.5)	(0.34, 4.50)
	2.00%	1.70%	0.36%	2.7669	299	(0.3, 4.5)	(0.30, 3.69)
						(0.9, 4.5)	(0.90, 4.50)
res.qdu.nr	2.19%	1.96%	0.40%	2.8329	300	(1.00, 4.58)	(1.00, 4.51)
	1.59%	1.48%	0.29%	3.0443	292	(1.00, 5.0)	(1.00, 4.99)
						(1.00, 5.0)	(1.00, 5.00)
res.qdu.ps	1.04%	2.36%	0.43%	3.0443	300	(0.6, 2.2)	(0.67, 2.11)
	1.79%	1.46%	0.14%	3.0449	289	(0.4, 2.2)	(0.50, 1.88)
						(0.7, 1.4)	(0.76, 1.34)

[1] For a key to modeling abbreviations, see Chapter 7, Exhibit 7.1.

[2] Unequal weighting effect defined as $1+[(n-1)/n]*CV^2$, where CV=coefficient of variation of weights.

[3] Number of proposed covariates on top line, and number finalized after modeling.

[4] There are six sets of bounds for each modeling step. Nominal bounds are used in defining maximum/minimum values for the GEM adjustment factors. The realized bound is the actual adjustment produced by the modeling. The set of three bounds listed for each step correspond to the high extreme values, the nonextreme values, and the low extreme values.

Table C2b 2002 Distribution of Weight Adjustment Factors and Weight Products (Model Group 2: Midwest)

	SDU wt	QDU design weight		sel.qdu.ps		res.qdu.nr		res.qdu.ps	
	1-9	duwght10	1-10	duwght11	1-11	duwght12	1-12	duwght13	1-13
Minimum	15	1.00	16	0.26	25	0.23	25	0.47	23
1%	119	1.00	123	0.60	119	1.00	123	0.90	134
5%	165	1.00	217	0.78	219	1.01	226	0.98	225
10%	283	1.00	412	0.85	375	1.03	388	0.99	385
25%	507	1.00	565	0.93	553	1.07	607	1.00	607
Median	614	1.04	753	1.00	763	1.12	848	1.00	853
75%	780	3.18	1,758	1.07	1,737	1.21	1,820	1.01	1,821
90%	1,199	6.14	4,021	1.17	4,097	1.32	4,793	1.01	4,813
95%	1,392	8.26	5,693	1.27	5,820	1.42	7,308	1.03	7,300
99%	1,891	13.29	10,974	1.61	11,104	1.72	14,206	1.11	14,244
Maximum	5,069	16.17	27,351	7.05	26,276	7.14	31,590	1.88	31,575
n	15,582	-	15,582	-	15,582	-	13,489	-	13,489
Mean	690	2.39	1,615	1.01	1,631	1.16	1,884	1.00	1,884
Max/Mean	7.34	-	16.94	-	16.11	-	16.77	-	16.76

[1] For a key to modeling abbreviations, see Chapter 7, Exhibit 7.1.

Model Group 2 Overview

Selected Questionnaire Dwelling Unit–Level Poststratification

A single adjustment of the initial set of covariates was required, such that the Indiana by race interaction combined Native American and Asian.

Respondent Questionnaire Dwelling Unit–Level Nonresponse

Insufficient sample for State-level race categories led to the creation of a combined Native American and Asian within Indiana, Iowa, Minnesota, Nebraska, North Dakota, and Ohio. Similarly, black, Native American, and Asian were combined in South Dakota. All other proposed factor effects were retained in full.

Respondent Questionnaire Dwelling Unit–Level Poststratification

For the final adjustment for the Midwest model group, all factors except State by race were kept at proposed levels. State by race was altered by combining Native American and Asian within each State.

Exhibit C2.1 Covariates for 2002 NSDUH Questionnaire Dwelling Unit Weights (sel.qdu.ps) Model Group 2: Midwest

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**66**	**66**	
Intercept	1	1	1	All levels present.
Group quarter	3	2	2	All levels present.
Race of Householder	6	5	5	All levels present.
Household Type	7	6	6	All levels present.
Household Size	1	1	1	All levels present.
Rent/Housing	5	4	4	All levels present.
Population Density	4	3	3	All levels present.
% Black	3	2	2	All levels present.
% Hispanic	35	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
State (count)	9	11	11	All levels present.
State (binary)	9	11	11	All levels present.
Quarter (count)	4	3	3	All levels present.
Quarter (binary)	4	3	3	All levels present.
Age Group	5	4	4	All levels present.
Race	5	4	4	All levels present.
Hispanicity	2	1	1	All levels present.
Gender	2	1	1	All levels present.
Two-Factor Effects		**163**	**162**	
Age x Race (3 levels)	5 x 3	8	8	All levels present.
Age x Hispanicity	5 x 2	4	4	All levels present.
Age x Gender	5 x 2	4	4	All levels present.
Race (3 levels) x Hispanicity	3 x 2	2	2	All levels present.
Race (3 levels) x Gender	3 x 2	2	2	All levels present.
Hispanicity x Gender	2 x 2	1	1	All levels present.
State x Age	12 x 5	44	44	All levels present.
State x Race	12 x 5	44	43	Coll (2,3) & (2,4); conv.
State x Gender	12 x 2	11	11	All levels present.
State x Hispanicity	12 x 2	11	11	All levels present.
% Black x % Owner-Occupied	3 x 3	4	4	All levels present.
% Black x Rent/Housing	3 x 5	8	8	All levels present.
% Hispanicity x % Owner-Occupied	3 x 3	4	3	All levels present.
% Hispanicity x Rent/housing	3 x 5	8	8	All levels present.
% Owner-Occupied x Rent/housing	3 x 5	8	8	All levels present.
Three-Factor Effects		**71**	**71**	
Race (3 levels) x Age x Gender	3 x 5 x 2	8	8	All levels present.
State/Region x Age x Gender	4 x 5 x 2	12	12	All levels present.
State/Region x Age x Hispanicity	4 x 5 x 2	12	12	All levels present.
State/Region x Age x Race (3 levels)	4 x 5 x 3	24	24	All levels present.
State/Region x Hispanicity x Gender	4 x 2 x 2	3	3	All levels present.
State/Region x Race (3 levels) x Hispanicity	4 x 3 x 2	6	6	All levels present.
State/Region x Race (3 levels) x Gender	4 x 3 x 2	6	6	All levels present.
Total		**300**	**299**	

Exhibit C2.2 Covariates for 2002 NSDUH Questionnaire Dwelling Unit Weights (res.qdu.nr) Model Group 2: Midwest

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**66**	**66**	
Intercept	1	1	1	All levels present.
Group quarter	3	2	2	All levels present.
Race of Householder	6	5	5	All levels present.
Household Type	7	6	6	All levels present.
Household Size	1	1	1	All levels present.
Rent/Housing	5	4	4	All levels present.
Population Density	4	3	3	All levels present.
% Black	3	2	2	All levels present.
% Hispanic	35	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
State (count)	9	11	11	All levels present.
State (binary)	9	11	11	All levels present.
Quarter (count)	4	3	3	All levels present.
Quarter (binary)	4	3	3	All levels present.
Age Group	5	4	4	All levels present.
Race	5	4	4	All levels present.
Hispanicity	2	1	1	All levels present.
Gender	2	1	1	All levels present.
Two-Factor Effects		**163**	**155**	
Age x Race (3 levels)	5 x 3	8	8	All levels present.
Age x Hispanicity	5 x 2	4	4	All levels present.
Age x Gender	5 x 2	4	4	All levels present.
Race (3 levels) x Hispanicity	3 x 2	2	2	All levels present.
Race (3 levels) x Gender	3 x 2	2	2	All levels present.
Hispanicity x Gender	2 x 2	1	1	All levels present.
State x Age	12 x 5	44	44	All levels present.
State x Race	12 x 5	44	36	Coll (2,3) & (2,4); conv. Repeat for States (3), (6), (8) (9), & (10); conv. Coll (11,2), (11,3), & (11,4); conv.
State x Gender	12 x 2	11	11	All levels present.
State x Hispanicity	12 x 2	11	11	All levels present.
% Black x % Owner-Occupied	3 x 3	4	4	All levels present.
% Black x Rent/Housing	3 x 5	8	8	All levels present.
% Hispanicity x % Owner-Occupied	3 x 3	4	3	All levels present.
% Hispanicity x Rent/housing	3 x 5	8	8	All levels present.
% Owner-Occupied x Rent/housing	3 x 5	8	8	All levels present.
Three-Factor Effects		**71**	**71**	
Race (3 levels) x Age x Gender	3 x 5 x 2	8	8	All levels present.
State/Region x Age x Gender	4 x 5 x 2	12	12	All levels present.
State/Region x Age x Hispanicity	4 x 5 x 2	12	12	All levels present.
State/Region x Age x Race (3 levels)	4 x 5 x 3	24	24	All levels present.
State/Region x Hispanicity x Gender	4 x 2 x 2	3	3	All levels present.
State/Region x Race (3 levels) x Hispanicity	4 x 3 x 2	6	6	All levels present.
State/Region x Race (3 levels) x Gender	4 x 3 x 2	6	6	All levels present.
Total		**300**	**292**	

Exhibit C2.3 Covariates for 2002 NSDUH Questionnaire Dwelling Unit Weights (res.qdu.ps) Model Group 2: Midwest

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**66**	**66**	
Intercept	1	1	1	All levels present.
Group quarter	3	2	2	All levels present.
Race of Householder	6	5	5	All levels present.
Household Type	7	6	6	All levels present.
Household Size	1	1	1	All levels present.
Rent/Housing	5	4	4	All levels present.
Population Density	4	3	3	All levels present.
% Black	3	2	2	All levels present.
% Hispanic	35	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
State (count)	9	11	11	All levels present.
State (binary)	9	11	11	All levels present.
Quarter (count)	4	3	3	All levels present.
Quarter (binary)	4	3	3	All levels present.
Age Group	5	4	4	All levels present.
Race	5	4	4	All levels present.
Hispanicity	2	1	1	All levels present.
Gender	2	1	1	All levels present.
Two-Factor Effects		**163**	**152**	
Age x Race (3 levels)	5 x 3	8	8	All levels present.
Age x Hispanicity	5 x 2	4	4	All levels present.
Age x Gender	5 x 2	4	4	All levels present.
Race (3 levels) x Hispanicity	3 x 2	2	2	All levels present.
Race (3 levels) x Gender	3 x 2	2	2	All levels present.
Hispanicity x Gender	2 x 2	1	1	All levels present.
State x Age	12 x 5	44	44	All levels present.
State x Race	12 x 5	44	33	Coll (1,3) & (1,4); conv. Repeat for all States.
State x Gender	12 x 2	11	11	All levels present.
State x Hispanicity	12 x 2	11	11	All levels present.
% Black x % Owner-Occupied	3 x 3	4	4	All levels present.
% Black x Rent/Housing	3 x 5	8	8	All levels present.
% Hispanicity x % Owner-Occupied	3 x 3	4	3	All levels present.
% Hispanicity x Rent/housing	3 x 5	8	8	All levels present.
% Owner-Occupied x Rent/housing	3 x 5	8	8	All levels present.
Three-Factor Effects		**71**	**71**	
Race (3 levels) x Age x Gender	3 x 5 x 2	8	8	All levels present.
State/Region x Age x Gender	4 x 5 x 2	12	12	All levels present.
State/Region x Age x Hispanicity	4 x 5 x 2	12	12	All levels present.
State/Region x Age x Race (3 levels)	4 x 5 x 3	24	24	All levels present.
State/Region x Hispanicity x Gender	4 x 2 x 2	3	3	All levels present.
State/Region x Race (3 levels) x Hispanicity	4 x 3 x 2	6	6	All levels present.
State/Region x Race (3 levels) x Gender	4 x 3 x 2	6	6	All levels present.
Total		**300**	**289**	

Appendix C3

Model Group 3: South

Alabama, Arkansas, Delaware, District of Columbia, Florida, Georgia, Kentucky, Louisiana, Maryland, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, Virginia, West Virginia

Table C3a 2002 QDU Weight GEM Modeling Summary (Model Group 3: South)

Modeling Step[1]	Extreme Weight Proportions			UWE[2]	# XVAR[3]	Bounds[4]	
	Unweighted	Weighted	Outwinsor			Nominal	Realized
sel.qdu.ps	1.40%	2.06%	0.41%	2.6703	339	(0.3, 3.3)	(0.44, 3.30)
	1.15%	1.68%	0.29%	2.6185	338	(0.3, 3.3)	(0.36, 3.26)
						(0.9, 3.3)	(0.90, 3.24)
res.qdu.nr	1.22%	1.70%	0.29%	2.6382	339	(1.00, 5.00)	(1.00, 5.00)
	1.04%	2.36%	0.43%	2.9754	323	(1.00, 5.00)	(1.00, 3.97)
						(1.00, 3.98)	(1.00, 3.58)
res.qdu.ps	1.04%	2.36%	0.43%	2.9754	339	(0.78, 3.00)	(0.80, 1.50)
	0.90%	2.10%	0.25%	2.9689	326	(0.67, 3.00)	(0.67, 1.50)
						(0.90, 1.24)	(0.79, 1.24)

[1] For a key to modeling abbreviations, see Chapter 7, Exhibit 7.1.

[2] Unequal weighting effect defined as $1+[(n-1)/n]*CV^2$, where CV=coefficient of variation of weights.

[3] Number of proposed covariates on top line, and number finalized after modeling.

[4] There are six sets of bounds for each modeling step. Nominal bounds are used in defining maximum/minimum values for the GEM adjustment factors. The realized bound is the actual adjustment produced by the modeling. The set of three bounds listed for each step correspond to the high extreme values, the nonextreme values, and the low extreme values.

Table C3b 2002 Distribution of Weight Adjustment Factors and Weight Products (Model Group 3: South)

	SDU wt	QDU design weight		sel.qdu.ps		res.qdu.nr		res.qdu.ps	
	1-9	duwght10	1-10	duwght11	1-11	duwght12	1-12	duwght13	1-13
Minimum	20	1.00	20	0.31	18	0.59	18	0.54	17
1%	75	1.00	85	0.64	87	1.00	91	0.94	90
5%	127	1.00	204	0.76	206	1.01	224	0.98	228
10%	283	1.00	435	0.82	412	1.02	443	0.99	439
25%	652	1.00	787	0.90	773	1.06	837	0.99	837
Median	921	1.08	1,215	0.99	1,209	1.11	1,301	1.00	1,303
75%	1,281	3.26	2,436	1.08	2,464	1.20	2,560	1.00	2,568
90%	1,668	6.41	5,966	1.20	5,965	1.33	6,861	1.01	6,843
95%	1,943	8.28	8,491	1.29	8,749	1.44	10,718	1.02	10,724
99%	2,614	13.22	14,624	1.63	14,634	1.73	18,089	1.10	18,134
Maximum	8,832	20.36	57,237	5.44	35,120	4.17	44,217	2.57	44,633
n	17,121	-	17,121	-	17,121	-	14,877	-	14,877
Mean	981	2.48	2,332	1.01	2,323	1.15	2,673	1.00	2,673
Max/Mean	9.00	-	24.54	-	15.12	-	15.54	-	16.69

[1] For a key to modeling abbreviations, see Chapter 7, Exhibit 7.1.

Model Group 3 Overview

Selected Questionnaire Dwelling Unit–Level Poststratification

All initially proposed covariates were maintained in this model except for the State by race variables District of Columbia by Native American and District of Columbia by Asian, which were collapsed.

Respondent Questionnaire Dwelling Unit–Level Nonresponse

The only changes from the proposed set of initial covariates involved the State by race interation. Levels Native American and Asian were collapsed in Alabama, Arkansas, District of Columbia, Georgia, Louisiana, Maryland, and South Carolina. Levels black, Native American, and Asian were collapsed in Mississippi, Tennessee, and West Virginia.

Respondent Questionnaire Dwelling Unit–Level Poststratification

As before, Native American and Asian variables were collapsed for the State by race interaction. States in which race was collapsed in this way include: Alabama, Arkansas, District of Columbia, Georgia, Kentucky, Louisiana, Maryland, Mississippi, South Carolina, Virginia and West Virginia. In addition, Tennessee State-level race was collapsed so that black, Native American, and Asian were combined.

Exhibit C3.1 Covariates for 2002 NSDUH Questionnaire Dwelling Unit Weights (sel.qdu.ps) Model Group 3: South

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**76**	**76**	
Intercept	1	1	1	All levels present.
Group quarter	3	2	2	All levels present.
Race of Householder	6	5	5	All levels present.
Household Type	7	6	6	All levels present.
Household Size	1	1	1	All levels present.
Rent/Housing	5	4	4	All levels present.
Population Density	4	3	3	All levels present.
% Black	3	2	2	All levels present.
% Hispanic	35	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
State (count)	17	16	16	All levels present.
State (binary)	17	16	16	All levels present.
Quarter (count)	4	3	3	All levels present.
Quarter (binary)	4	3	3	All levels present.
Age Group	5	4	4	All levels present.
Race	5	4	4	All levels present.
Hispanicity	2	1	1	All levels present.
Gender	2	1	1	All levels present.
Two-Factor Effects		**213**	**212**	
Age x Race (3 levels)	5 x 3	8	8	All levels present.
Age x Hispanicity	5 x 2	4	4	All levels present.
Age x Gender	5 x 2	4	4	All levels present.
Race (3 levels) x Hispanicity	3 x 2	2	2	All levels present.
Race (3 levels) x Gender	3 x 2	2	2	All levels present.
Hispanicity x Gender	2 x 2	1	1	All levels present.
State x Age	17 x 5	64	64	All levels present.
State x Race	17 x 5	64	63	Coll (4,3) & (4,4); conv.
State x Gender	17 x 2	16	16	All levels present.
State x Hispanicity	17 x 2	16	16	All levels present.
% Black x % Owner-Occupied	3 x 3	4	4	All levels present.
% Black x Rent/Housing	3 x 5	8	8	All levels present.
% Hispanicity x % Owner-Occupied	3 x 3	4	3	All levels present.
% Hispanicity x Rent/housing	3 x 5	8	8	All levels present.
% Owner-Occupied x Rent/housing	3 x 5	8	8	All levels present.
Three-Factor Effects		**71**	**71**	
Race (3 levels) x Age x Gender	3 x 5 x 2	8	8	All levels present.
State/Region x Age x Gender	3 x 5 x 2	8	8	All levels present.
State/Region x Age x Hispanicity	3 x 5 x 2	8	8	All levels present.
State/Region x Age x Race (3 levels)	3 x 5 x 3	16	16	All levels present.
State/Region x Hispanicity x Gender	3 x 2 x 2	2	2	All levels present.
State/Region x Race (3 levels) x Hispanicity	3 x 3 x 2	4	4	All levels present.
State/Region x Race (3 levels) x Gender	3 x 3 x 2	4	4	All levels present.
Total		**339**	**338**	

Exhibit C3.2 Covariates for 2002 NSDUH Questionnaire Dwelling Unit Weights (res.qdu.nr) Model Group 3: South

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**75**	**75**	
Intercept	1	1	1	All levels present.
Group quarter	3	2	2	All levels present.
Race of Householder	6	5	5	All levels present.
Household Type	7	6	6	All levels present.
Household Size	1	1	1	All levels present.
Rent/Housing	5	4	4	All levels present.
Population Density	4	3	3	All levels present.
% Black	3	2	2	All levels present.
% Hispanic	35	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
State (count)	17	16	16	All levels present.
State (binary)	17	16	16	All levels present.
Quarter (count)	4	3	3	All levels present.
Quarter (binary)	4	3	3	All levels present.
Age Group	5	4	4	All levels present.
Race	5	4	4	All levels present.
Hispanicity	2	1	1	All levels present.
Gender	2	1	1	All levels present.
Two-Factor Effects		**213**	**197**	
Age x Race (3 levels)	5 x 3	8	8	All levels present.
Age x Hispanicity	5 x 2	4	4	All levels present.
Age x Gender	5 x 2	4	4	All levels present.
Race (3 levels) x Hispanicity	3 x 2	2	2	All levels present.
Race (3 levels) x Gender	3 x 2	2	2	All levels present.
Hispanicity x Gender	2 x 2	1	1	All levels present.
State x Age	17 x 5	64	64	All levels present.
State x Race	17 x 5	64	48	Coll. (1,3) & (1,4); conv. Repeat for States (2), (4), (6), (8), (9), (13), & (16). Coll. (7,3), (7,4), & (7,5); conv. Repeat for States (10), (14), & (17)
State x Gender	17 x 2	16	16	All levels present.
State x Hispanicity	17 x 2	16	16	All levels present.
% Black x % Owner-Occupied	3 x 3	4	4	All levels present.
% Black x Rent/Housing	3 x 5	8	8	All levels present.
% Hispanicity x % Owner-Occupied	3 x 3	4	4	All levels present.
% Hispanicity x Rent/housing	3 x 5	8	8	All levels present.
% Owner-Occupied x Rent/housing	3 x 5	8	8	All levels present.
Three-Factor Effects		**71**	**71**	
Race (3 levels) x Age x Gender	3 x 5 x 2	8	8	All levels present.
State/Region x Age x Gender	3 x 5 x 2	8	8	All levels present.
State/Region x Age x Hispanicity	3 x 5 x 2	8	8	All levels present.
State/Region x Age x Race (3 levels)	3 x 5 x 3	16	16	All levels present.
State/Region x Hispanicity x Gender	3 x 2 x 2	2	2	All levels present.
State/Region x Race (3 levels) x Hispanicity	3 x 3 x 2	4	4	All levels present.
State/Region x Race (3 levels) x Gender	3 x 3 x 2	4	4	All levels present.
Total		**339**	**323**	

Exhibit C3.3 Covariates for 2002 NSDUH Questionnaire Dwelling Unit Weights (res.qdu.ps) Model Group 3: South

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**76**	**76**	
Intercept	1	1	1	All levels present.
Group quarter	3	2	2	All levels present.
Race of Householder	6	5	5	All levels present.
Household Type	7	6	6	All levels present.
Household Size	1	1	1	All levels present.
Rent/Housing	5	4	4	All levels present.
Population Density	4	3	3	All levels present.
% Black	3	2	2	All levels present.
% Hispanic	35	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
State (count)	17	16	16	All levels present.
State (binary)	17	16	16	All levels present.
Quarter (count)	4	3	3	All levels present.
Quarter (binary)	4	3	3	All levels present.
Age Group	5	4	4	All levels present.
Race	5	4	4	All levels present.
Hispanicity	2	1	1	All levels present.
Gender	2	1	1	All levels present.
Two-Factor Effects		**213**	**200**	
Age x Race (3 levels)	5 x 3	8	8	All levels present.
Age x Hispanicity	5 x 2	4	4	All levels present.
Age x Gender	5 x 2	4	4	All levels present.
Race (3 levels) x Hispanicity	3 x 2	2	2	All levels present.
Race (3 levels) x Gender	3 x 2	2	2	All levels present.
Hispanicity x Gender	2 x 2	1	1	All levels present.
State x Age	17 x 5	64	64	All levels present.
State x Race	17 x 5	64	51	Coll. (1,3) & (1,4); conv. Repeat for states (2), (4), (6), (7), (8), (9), (10), (13), (16), & (17). Coll. (14,3), (14,4), & (14,5); conv
State x Gender	17 x 2	16	16	All levels present.
State x Hispanicity	17 x 2	16	16	All levels present.
% Black x % Owner-Occupied	3 x 3	4	4	All levels present.
% Black x Rent/Housing	3 x 5	8	8	All levels present.
% Hispanicity x % Owner-Occupied	3 x 3	4	4	All levels present.
% Hispanicity x Rent/housing	3 x 5	8	8	All levels present.
% Owner-Occupied x Rent/housing	3 x 5	8	8	All levels present.
Three-Factor Effects		**71**	**71**	
Race (3 levels) x Age x Gender	3 x 5 x 2	8	8	All levels present.
State/Region x Age x Gender	3 x 5 x 2	8	8	All levels present.
State/Region x Age x Hispanicity	3 x 5 x 2	8	8	All levels present.
State/Region x Age x Race (3 levels)	3 x 5 x 3	16	16	All levels present.
State/Region x Hispanicity x Gender	3 x 2 x 2	2	2	All levels present.
State/Region x Race (3 levels) x Hispanicity	3 x 3 x 2	4	4	All levels present.
State/Region x Race (3 levels) x Gender	3 x 3 x 2	4	4	All levels present.
Total		**339**	**323**	

Appendix C4

Model Group 4: West

Alaska, Arizona, California, Colorado, Idaho, Hawaii, Montana, Nevada, New Mexico, Oregon, Utah, Washington, Wyoming

Table C4a 2002 QDU Weight GEM Modeling Summary (Model Group 4: West)

Modeling Step[1]	Extreme Weight Proportions			UWE[2]	# XVAR[3]	Bounds[4]	
	Unweighted	Weighted	Outwinsor			Nominal	Realized
sel.qdu.ps	1.36%	2.23%	0.35%	3.1733	270	(0.2, 3.40)	(0.20, 3.40)
	1.45%	2.25%	0.49%	3.2800	268	(0.2, 3.40)	(0.20, 3.40)
						(0.9, 3.40)	(0.90, 3.39)
res.qdu.nr	1.67%	4.03%	0.64%	3.2619	270	(1.00, 3.44)	(1.00, 3.38)
	1.57%	5.30%	0.83%	3.8436	266	(1.00, 5.00)	(1.00, 5.00)
						(1.00, 5.00)	(1.00, 5.00)
res.qdu.ps	1.57%	5.30%	0.83%	3.8437	270	(0.84, 2.60)	(0.92, 2.60)
	1.71%	5.20%	0.38%	3.8395	267	(0.84, 2.60)	(0.89, 1.75)
						(0.88, 1.26)	(0.89, 1.13)

[1] For a key to modeling abbreviations, see Chapter 7, Exhibit 7.1.

[2] Unequal weighting effect defined as $1+[(n-1)/n]*CV^2$, where CV=coefficient of variation of weights.

[3] Number of proposed covariates on top line, and number finalized after modeling.

[4] There are six sets of bounds for each modeling step. Nominal bounds are used in defining maximum/minimum values for the GEM adjustment factors. The realized bound is the actual adjustment produced by the modeling. The set of three bounds listed for each step correspond to the high extreme values, the nonextreme values, and the low extreme values.

Table C4b 2002 Distribution of Weight Adjustment Factors and Weight Products (Model Group 4: West)

	SDU wt	QDU design weight		sel.qdu.ps		res.qdu.nr		res.qdu.ps	
	1-9	duwght10	1-10	duwght11	1-11	duwght12	1-12	duwght13	1-13
Minimum	14	1.00	14	0.20	13	0.37	13	0.60	13
1%	84	1.00	94	0.57	79	1.00	88	0.92	88
5%	117	1.00	131	0.76	128	1.01	141	0.97	141
10%	140	1.00	172	0.82	173	1.02	187	0.98	187
25%	262	1.00	399	0.90	395	1.05	424	0.99	424
Median	679	1.08	1,067	0.98	1,055	1.11	1,134	1.00	1,134
75%	1,447	3.10	2,127	1.09	2,131	1.21	2,299	1.01	2,299
90%	1,888	5.73	5,358	1.19	5,375	1.36	5,846	1.02	5,846
95%	2,208	7.73	8,033	1.28	8,050	1.50	9,692	1.03	9,692
99%	2,865	12.88	14,771	1.64	15,498	1.90	21,015	1.09	21,015
Maximum	6,085	15.76	35,038	3.40	34,950	5.00	52,370	2.07	52,370
n	11,547	-	11,547	-	11,547	-	9,998	-	9,998
Mean	892	2.36	2,036	1.00	2,048	1.16	2,365	1.00	2,365
Max/Mean	6.82	-	17.21	-	17.07	-	22.15	-	22.47

[1] For a key to modeling abbreviations, see Chapter 7, Exhibit 7.1.

Model Group 4 Overview

Selected Questionnaire Dwelling Unit–Level Poststratification

The set of variables included in the modeling of the selected QDU-level poststratification differed from the originally proposed set only in that two segment-level effects were collapsed with the reference due to zero sample size. These factor effects were levels of segment percent black by segment percent owner-occupied dwelling units and segment percent black by median rent/housing value.

Respondent Questionnaire Dwelling Unit–Level Nonresponse

As in the preceding Selected Questionnaire Dwelling Unit-Level Poststratification step, the segment-level factor effects percent black by percent owner-occupied dwelling units and percent black by median rent/housing value had levels removed due to zero sample. In addition, the State by race collapsed Arizona black and Arizona Native American, as well as Montana Asian with Montana multi-race.

Respondent Questionnaire Dwelling Unit–Level Poststratification

This step removed the same segment level variables due to zero counts as the preceding steps, except that State by race collapsed Alaska black and Alaska Native American. All other factor effects were retained at proposed levels.

Exhibit C4.1 Covariates for 2002 NSDUH Questionnaire Dwelling Unit Weights (sel.qdu.ps) Model Group 4: West

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**68**	**68**	
Intercept	1	1	1	All levels present.
Group quarter	3	2	2	All levels present.
Race of Householder	6	5	5	All levels present.
Household Type	7	6	6	All levels present.
Household Size	1	1	1	All levels present.
Rent/Housing	5	4	4	All levels present.
Population Density	4	3	3	All levels present.
% Black	3	2	2	All levels present.
% Hispanic	35	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
State (count)	13	12	12	All levels present.
State (binary)	13	12	12	All levels present.
Quarter (count)	4	3	3	All levels present.
Quarter (binary)	4	3	3	All levels present.
Age Group	5	4	4	All levels present.
Race	5	4	4	All levels present.
Hispanicity	2	1	1	All levels present.
Gender	2	1	1	All levels present.
Two-Factor Effects		**173**	**171**	
Age x Race (3 levels)	5 x 3	8	8	All levels present.
Age x Hispanicity	5 x 2	4	4	All levels present.
Age x Gender	5 x 2	4	4	All levels present.
Race (3 levels) x Hispanicity	3 x 2	2	2	All levels present.
Race (3 levels) x Gender	3 x 2	2	2	All levels present.
Hispanicity x Gender	2 x 2	1	1	All levels present.
State x Age	13 x 5	48	48	All levels present.
State x Race	13 x 5	48	48	Coll. (1,3) & (2,3); zero cnts.
State x Gender	13 x 2	12	12	Coll. (1,1) & (2,1); zero cnts.
State x Hispanicity	13 x 2	12	12	All levels present.
% Black x % Owner-Occupied	3 x 3	4	4	All levels present.
% Black x Rent/Housing	3 x 5	8	8	All levels present.
% Hispanicity x % Owner-Occupied	3 x 3	4	4	All levels present.
% Hispanicity x Rent/housing	3 x 5	8	8	All levels present.
% Owner-Occupied x Rent/housing	3 x 5	8	8	All levels present.
Three-Factor Effects		**71**	**71**	
Race (3 levels) x Age x Gender	3 x 5 x 2	8	8	All levels present.
State/Region x Age x Gender	2 x 5 x 2	4	4	All levels present.
State/Region x Age x Hispanicity	2 x 5 x 2	4	4	All levels present.
State/Region x Age x Race (3 levels)	2 x 5 x 3	8	8	All levels present.
State/Region x Hispanicity x Gender	2 x 2 x 2	1	1	All levels present.
State/Region x Race (3 levels) x Hispanicity	2 x 3 x 2	2	2	All levels present.
State/Region x Race (3 levels) x Gender	2 x 3 x 2	2	2	All levels present.
Total		**270**	**268**	

Exhibit C4.2 Covariates for 2002 NSDUH Questionnaire Dwelling Unit Weights (res.qdu.nr) Model Group 4: West

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**68**	**68**	
Intercept	1	1	1	All levels present.
Group quarter	3	2	2	All levels present.
Race of Householder	6	5	5	All levels present.
Household Type	7	6	6	All levels present.
Household Size	1	1	1	All levels present.
Rent/Housing	5	4	4	All levels present.
Population Density	4	3	3	All levels present.
% Black	3	2	2	All levels present.
% Hispanic	35	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
State (count)	13	12	12	All levels present.
State (binary)	13	12	12	All levels present.
Quarter (count)	4	3	3	All levels present.
Quarter (binary)	4	3	3	All levels present.
Age Group	5	4	4	All levels present.
Race	5	4	4	All levels present.
Hispanicity	2	1	1	All levels present.
Gender	2	1	1	All levels present.
Two-Factor Effects		**173**	**169**	
Age x Race (3 levels)	5 x 3	8	8	All levels present.
Age x Hispanicity	5 x 2	4	4	All levels present.
Age x Gender	5 x 2	4	4	All levels present.
Race (3 levels) x Hispanicity	3 x 2	2	2	All levels present.
Race (3 levels) x Gender	3 x 2	2	2	All levels present.
Hispanicity x Gender	2 x 2	1	1	All levels present.
State x Age	13 x 5	48	48	All levels present.
State x Race	13 x 5	48	46	Coll. (2,2) & (2,3), (7,4) & (7,5); conv.
State x Gender	13 x 2	12	12	Coll. (1,1) & (2,1); zero cnts.
State x Hispanicity	13 x 2	12	12	All levels present.
% Black x % Owner-Occupied	3 x 3	4	3	Coll. (1,3) & (2,3); zero cnts.
% Black x Rent/Housing	3 x 5	8	7	Coll. (1,1) & (2,1); zero cnts.
% Hispanicity x % Owner-Occupied	3 x 3	4	4	All levels present.
% Hispanicity x Rent/housing	3 x 5	8	8	All levels present.
% Owner-Occupied x Rent/housing	3 x 5	8	8	All levels present.
Three-Factor Effects		**29**	**29**	
Race (3 levels) x Age x Gender	3 x 5 x 2	8	8	All levels present.
State/Region x Age x Gender	2 x 5 x 2	4	4	All levels present.
State/Region x Age x Hispanicity	2 x 5 x 2	4	4	All levels present.
State/Region x Age x Race (3 levels)	2 x 5 x 3	8	8	All levels present.
State/Region x Hispanicity x Gender	2 x 2 x 2	1	1	All levels present.
State/Region x Race (3 levels) x Hispanicity	2 x 3 x 2	2	2	All levels present.
State/Region x Race (3 levels) x Gender	2 x 3 x 2	2	2	All levels present.
Total		**270**	**266**	

Exhibit C4.3 Covariates for 2002 NSDUH Questionnaire Dwelling Unit Weights (res.qdu.ps) Model Group 4: West

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**68**	**68**	
Intercept	1	1	1	All levels present.
Group quarter	3	2	2	All levels present.
Race of Householder	6	5	5	All levels present.
Household Type	7	6	6	All levels present.
Household Size	1	1	1	All levels present.
Rent/Housing	5	4	4	All levels present.
Population Density	4	3	3	All levels present.
% Black	3	2	2	All levels present.
% Hispanic	35	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
State (count)	13	12	12	All levels present.
State (binary)	13	12	12	All levels present.
Quarter (count)	4	3	3	All levels present.
Quarter (binary)	4	3	3	All levels present.
Age Group	5	4	4	All levels present.
Race	5	4	4	All levels present.
Hispanicity	2	1	1	All levels present.
Gender	2	1	1	All levels present.
Two-Factor Effects		**173**	**170**	
Age x Race (3 levels)	5 x 3	8	8	All levels present.
Age x Hispanicity	5 x 2	4	4	All levels present.
Age x Gender	5 x 2	4	4	All levels present.
Race (3 levels) x Hispanicity	3 x 2	2	2	All levels present.
Race (3 levels) x Gender	3 x 2	2	2	All levels present.
Hispanicity x Gender	2 x 2	1	1	All levels present.
State x Age	13 x 5	48	48	All levels present.
State x Race	13 x 5	48	47	Coll. (1,2) & (1,3); conv..
State x Gender	13 x 2	12	12	Coll. (1,1) & (2,1); zero cnts.
State x Hispanicity	13 x 2	12	12	All levels present.
% Black x % Owner-Occupied	3 x 3	4	3	Coll. (1,3) & (2,3); zero cnts.
% Black x Rent/Housing	3 x 5	8	7	Coll. (1,1) & (2,1); zero cnts.
% Hispanicity x % Owner-Occupied	3 x 3	4	4	All levels present.
% Hispanicity x Rent/housing	3 x 5	8	8	All levels present.
% Owner-Occupied x Rent/housing	3 x 5	8	8	All levels present.
Three-Factor Effects		**29**	**29**	
Race (3 levels) x Age x Gender	3 x 5 x 2	8	8	All levels present.
State/Region x Age x Gender	2 x 5 x 2	4	4	All levels present.
State/Region x Age x Hispanicity	2 x 5 x 2	4	4	All levels present.
State/Region x Age x Race (3 levels)	2 x 5 x 3	8	8	All levels present.
State/Region x Hispanicity x Gender	2 x 2 x 2	1	1	All levels present.
State/Region x Race (3 levels) x Hispanicity	2 x 3 x 2	2	2	All levels present.
State/Region x Race (3 levels) x Gender	2 x 3 x 2	2	2	All levels present.
Total		**270**	**267**	

Appendix D

Evaluation of Calibration Weights: Questionnaire Dwelling Unit–Level Response Rates

Table D.1 2002 NSDUH QDU-Level Response Rates

Domain	Selected QDU	Respondent QDU	Interview Response Rate [1]
Total	55,686	48,088	80.60%
Census Region			
Northeast	11,436	9,724	77.28%
South	17,121	14,877	81.73%
Midwest	15,582	13,489	81.87%
West	11,547	9,998	80.23%
Quarter			
Quarter 1	13,703	12,014	83.09%
Quarter 2	13,579	11,692	80.28%
Quarter 3	15,149	13,060	79.85%
Quarter 4	13,255	11,322	79.20%
Household Type			
12-17, 18-25, 26+	4,959	4,539	90.78%
12-17, 18-25	131	116	84.21%
12-17, 26+	17,089	15,573	91.01%
18-25, 26+	10,878	9,239	84.52%
12-17	78	72	92.06%
18-25	7,062	6,396	90.91%
26+	15,489	12,153	76.52%
Race of Householder			
Hispanic white	5,798	5,087	84.30%
Hispanic black	163	144	83.92%
Hispanic other	351	308	86.24%
Non-Hispanic white	39,367	33,802	79.68%
Non-Hispanic black	6,616	5,846	84.01%
Non-Hispanic other	3,391	2,901	78.44%
% Hispanic in Segment			
50-100%	2,817	2,458	83.30%
10-50%	9,109	7,856	80.19%
<10%	43,760	37,774	80.49%
% Black in Segment			
50-100%	4,429	3,898	83.14%
10-50%	7,935	6,896	81.88%
<10%	43,322	37,294	80.08%
% Owner-Occupied DUs in Segment			
50-100%	41,619	35,824	80.22%
10-50%	10,720	9,319	81.39%
<10%	3,347	2,945	83.02%
Combined Median Rent/Housing Value			
1st Quintile	10,526	9,184	81.42%
2nd Quintile	11,363	9,846	81.63%
3rd Quintile	11,690	10,105	80.32%
4th Quintile	10,892	9,365	79.74%
5th Quintile	11,215	9,588	80.32%
Population Density			
Large MSA	19,841	16,872	79.22%
Medium-Small MSA	20,648	17,883	80.64%
Non-MSA, Urban	6,771	6,001	84.81%
Non-MSA, Rural	8,426	7,332	82.06%
Group Quarters			
Group	1,046	1,013	97.58%
Nongroup	54,640	47,075	80.44%
Household Size			
One	7,101	5,884	77.65%
Two	20,520	17,232	79.01%
Three	15,742	13,802	84.80%
Four or more	12,323	11,170	88.75%

[1] The weight used for calculating the response rate includes SDU-level and QDU-level design weights, SDU nonresponse and poststratification adjustments, and selected QDU poststratification adjustment. This weight is the product of YR02WT1*...*YR02WT9*DU02WT10*DU02WT11

Appendix E

Evaluation of Calibration Weights: Questionnaire Dwelling Unit–Level Proportions of Extreme Values and Outwinsors

Table E.1 2002 NSDUH Selected QDU-Level Proportions of Extreme Values and Outwinsors

Domain	*n*	Screener DU-Level Weights (SDUWT: YR02WT1*...*YR02WT9)			Before sel.qdu.ps[1] (SDUWT*DU02WT10)			After sel.qdu.ps[1] (SDUWT*DU02WT10*DU02WT11)		
		Unweighted	Weighted[2]	Outwinsor[3]	Unweighted	Weighted[2]	Outwinsor[3]	Unweighted	Weighted[2]	Outwinsor[3]
Total	55,686	2.30%	3.61%	0.66%	1.73%	2.04%	0.37%	1.56%	1.92%	0.40%
Census Region										
Northeast	11,436	2.47%	4.80%	1.16%	2.01%	2.31%	0.50%	1.61%	2.28%	0.57%
South	17,121	1.65%	2.69%	0.55%	1.40%	2.06%	0.41%	1.15%	1.68%	0.29%
Midwest	15,582	2.88%	4.01%	0.50%	2.17%	1.61%	0.24%	2.04%	1.70%	0.36%
West	11,547	2.35%	3.73%	0.58%	1.36%	2.23%	0.35%	1.45%	2.25%	0.49%
Quarter										
Quarter 1	13,703	2.34%	3.55%	0.70%	1.77%	2.20%	0.36%	1.50%	1.70%	0.42%
Quarter 2	13,579	2.21%	3.95%	0.91%	1.69%	2.88%	0.69%	1.86%	2.74%	0.58%
Quarter 3	15,149	1.79%	2.64%	0.40%	1.33%	1.26%	0.21%	1.14%	1.17%	0.29%
Quarter 4	13,255	2.96%	4.29%	0.63%	2.20%	1.85%	0.25%	1.79%	2.07%	0.33%
Household Type										
12-17, 18-25, 26+	4,959	2.00%	3.58%	0.73%	2.00%	3.58%	0.73%	1.98%	4.66%	1.23%
12-17, 18-25	131	0.76%	2.45%	0.19%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
12-17, 26+	17,089	2.14%	3.48%	0.68%	2.09%	3.42%	0.68%	2.01%	4.10%	1.00%
18-25, 26+	10,878	2.31%	3.48%	0.62%	2.14%	3.23%	0.60%	1.98%	3.99%	1.02%
12-17	78	1.28%	1.81%	0.62%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
18-25	7,062	2.93%	4.52%	0.84%	2.66%	4.41%	0.79%	2.15%	3.77%	0.75%
26+	15,489	2.31%	3.47%	0.56%	0.56%	1.30%	0.23%	0.38%	0.87%	0.12%
Race of Householder										
Hispanic white	5,798	2.88%	3.71%	0.49%	2.10%	1.99%	0.25%	1.91%	2.07%	0.36%
Hispanic black	163	38.04%	58.29%	19.77%	31.90%	37.95%	11.25%	23.31%	28.16%	12.78%
Hispanic other	351	17.09%	25.54%	5.42%	12.25%	19.99%	4.12%	7.98%	10.26%	3.65%
Non-Hispanic white	39,367	1.08%	1.47%	0.21%	0.78%	1.02%	0.12%	0.62%	0.85%	0.12%
Non-Hispanic black	6,616	3.45%	6.13%	1.14%	2.89%	4.34%	1.08%	3.39%	4.58%	0.94%
Non-Hispanic other	3,391	10.03%	15.02%	2.82%	7.34%	6.99%	1.46%	6.55%	8.20%	2.01%
% Hispanic in Segment										
50-100%	2,817	2.24%	4.13%	0.98%	1.67%	2.52%	0.58%	1.70%	3.08%	0.86%
10-50%	9,109	2.57%	4.32%	0.84%	1.91%	2.32%	0.45%	2.07%	2.62%	0.70%
<10%	43,760	2.25%	3.36%	0.57%	1.70%	1.93%	0.34%	1.44%	1.66%	0.29%

Table E.1 2002 NSDUH Selected QDU-Level Proportions of Extreme Values and Outwinsors (continued)

Domain	*n*	Screener DU-Level Weights (SDUWT: YR02WT1*...*YR02WT9)			Before sel.qdu.ps[1] (SDUWT*DU02WT10)			After sel.qdu.ps[1] (SDUWT*DU02WT10*DU02WT11)		
		Unweighted	Weighted[2]	Outwinsor[3]	Unweighted	Weighted[2]	Outwinsor[3]	Unweighted	Weighted[2]	Outwinsor[3]
% Black in Segment										
50-100%	4,429	2.96%	5.33%	1.08%	2.53%	4.86%	1.33%	3.57%	5.39%	1.15%
10-50%	7,935	2.94%	5.01%	1.10%	2.57%	3.03%	0.59%	2.07%	2.59%	0.73%
<10%	43,322	2.12%	3.10%	0.51%	1.50%	1.56%	0.24%	1.26%	1.43%	0.26%
% Owner-Occupied DUs in Segment										
50-100%	41,619	2.02%	2.92%	0.48%	1.49%	1.81%	0.32%	1.26%	1.66%	0.32%
10-50%	10,720	3.01%	5.38%	1.19%	2.35%	2.66%	0.55%	2.35%	2.61%	0.64%
<10%	3,347	3.53%	5.93%	1.04%	2.75%	2.99%	0.55%	2.66%	3.13%	0.77%
Combined Median Rent/Housing Value										
1st Quintile	10,526	2.68%	3.62%	0.51%	1.79%	1.58%	0.28%	1.38%	1.36%	0.30%
2nd Quintile	11,363	2.15%	3.13%	0.61%	1.43%	1.20%	0.22%	1.27%	1.17%	0.20%
3rd Quintile	11,690	2.37%	3.89%	0.71%	1.97%	2.58%	0.41%	1.83%	2.59%	0.40%
4th Quintile	10,892	2.35%	3.57%	0.63%	1.90%	2.21%	0.52%	1.66%	2.23%	0.45%
5th Quintile	11,215	2.00%	3.75%	0.77%	1.57%	2.36%	0.39%	1.63%	1.96%	0.59%
Population Density										
Large MSA	19,841	2.57%	4.21%	0.74%	2.07%	2.43%	0.37%	1.89%	2.27%	0.46%
Medium-Small MSA	20,648	2.37%	3.73%	0.77%	1.68%	2.01%	0.47%	1.60%	2.03%	0.47%
Non-MSA, Urban	6,771	1.98%	2.49%	0.46%	1.57%	1.51%	0.29%	1.11%	1.26%	0.23%
Non-MSA, Rural	8,426	1.78%	2.02%	0.21%	1.20%	1.17%	0.19%	1.03%	0.95%	0.16%
Group Quarters										
Group	1,046	3.15%	4.02%	0.62%	3.15%	3.25%	0.40%	2.87%	4.00%	0.81%
Nongroup	54,640	2.29%	3.60%	0.66%	1.70%	2.03%	0.37%	1.53%	1.90%	0.40%
Household Size										
One	7,101	1.96%	2.67%	0.35%	1.25%	1.11%	0.32%	0.89%	0.90%	0.11%
Two	20,520	2.15%	3.39%	0.65%	1.40%	1.94%	0.30%	1.18%	1.51%	0.29%
Three	15,742	2.37%	3.48%	0.60%	1.98%	2.93%	0.49%	1.89%	3.13%	0.82%
Four or more	12,323	2.67%	4.56%	0.88%	2.24%	3.58%	0.69%	2.14%	4.51%	1.02%

[1] Sel = selected, QDU = questionnaire dwelling unit, PS = poststratification adjustment.
[2] Weighted extreme value proportion: $100 * \sum_k w_{ek}/\sum_k w_k$, where w_{ek} denotes the weight for extreme values and w_k denotes the weight for both extreme values and non-extreme values.
[3] Outwinsor weight proportion: $100 * \sum_k (w_{ek}-b_k)/\sum_k w_k$, where b_k denotes the winsorized weight.

Table E.2 2002 NSDUH Respondent QDU-Level Proportions of Extreme Values and Outwinsors

Domain	*n*	Before res.qdu.nr[1] (SDUWT*DU02WT10*DU02WT11)			After res.qdu.nr[1] (SDUWT*DU02WT10*...*DU02WT12)			Final Weight: After res.qdu.ps[1] (SDUWT*DU02WT10*...*DU02WT13)		
		Unweighted	Weighted[2]	Outwinsor[3]	Unweighted	Weighted[2]	Outwinsor[3]	Unweighted	Weighted[2]	Outwinsor[3]
Total	48,088	1.68%	2.40%	0.46%	1.39%	2.88%	0.54%	1.47%	2.68%	0.31%
Census Region										
Northeast	9,724	1.70%	2.45%	0.67%	1.49%	2.82%	0.75%	1.65%	2.41%	0.55%
South	14,877	1.22%	1.70%	0.29%	1.04%	2.36%	0.43%	0.90%	2.10%	0.25%
Midwest	13,489	2.19%	1.96%	0.40%	1.59%	1.48%	0.29%	1.79%	1.46%	0.14%
West	9,998	1.67%	4.03%	0.64%	1.57%	5.30%	0.83%	1.71%	5.20%	0.38%
Quarter										
Quarter 1	12,014	1.66%	2.90%	0.59%	1.26%	2.42%	0.56%	1.35%	2.44%	0.35%
Quarter 2	11,692	1.95%	2.97%	0.53%	1.53%	3.10%	0.68%	1.56%	2.78%	0.39%
Quarter 3	13,060	1.26%	1.26%	0.34%	1.03%	2.14%	0.34%	1.17%	2.01%	0.20%
Quarter 4	11,322	1.93%	2.44%	0.38%	1.81%	3.84%	0.60%	1.85%	3.49%	0.31%
Household Type										
12-17, 18-25, 26+	4,539	1.81%	4.48%	1.18%	1.83%	4.72%	1.41%	1.81%	4.51%	1.03%
12-17, 18-25	116	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
12-17, 26+	15,573	2.07%	4.05%	0.98%	1.55%	3.68%	0.97%	1.64%	3.53%	0.65%
18-25, 26+	9,239	2.09%	4.37%	1.12%	1.87%	5.08%	1.61%	1.90%	4.53%	0.98%
12-17	72	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	1.39%	3.48%	0.09%
18-25	6,396	2.17%	3.67%	0.76%	1.34%	2.70%	0.60%	1.59%	2.87%	0.45%
26+	12,153	0.60%	1.42%	0.15%	0.72%	2.30%	0.24%	0.74%	2.11%	0.09%
Race of Householder										
Hispanic white	5,087	2.06%	2.37%	0.42%	1.47%	1.97%	0.51%	1.97%	2.07%	0.29%
Hispanic black	144	21.53%	28.57%	13.79%	20.83%	27.11%	10.25%	24.31%	29.51%	10.66%
Hispanic other	308	9.09%	13.89%	4.45%	7.47%	9.23%	2.24%	6.49%	9.14%	1.89%
Non-Hispanic white	33,802	0.73%	1.37%	0.17%	0.74%	2.24%	0.27%	0.69%	1.97%	0.10%
Non-Hispanic black	5,846	3.44%	4.52%	0.93%	2.31%	3.82%	0.91%	2.34%	3.62%	0.56%
Non-Hispanic other	2,901	6.79%	9.00%	2.04%	5.45%	9.05%	2.77%	6.31%	9.18%	1.90%
% Hispanic in Segment										
50-100%	2,458	1.91%	3.67%	1.08%	1.59%	2.32%	0.57%	1.63%	2.32%	0.53%
10-50%	7,856	2.19%	3.04%	0.72%	1.79%	3.27%	0.83%	1.99%	3.02%	0.52%
<10%	37,774	1.56%	2.14%	0.35%	1.30%	2.82%	0.47%	1.35%	2.62%	0.24%

Table E.2 2002 NSDUH Respondent QDU-Level Proportions of Extreme Values and Outwinsors (continued)

Domain	*n*	Before res.qdu.nr[1] (SDUWT*DU02WT10*DU02WT11)			After res.qdu.nr[1] (SDUWT*DU02WT10*...*DU02WT12)			Final Weight: After res.qdu.ps[1] (SDUWT*DU02WT10*...*DU02WT13)		
		Unweighted	Weighted[2]	Outwinsor[3]	Unweighted	Weighted[2]	Outwinsor[3]	Unweighted	Weighted[2]	Outwinsor[3]
% Black in Segment										
50-100%	3,898	3.62%	4.91%	0.91%	2.23%	4.67%	1.07%	2.18%	4.22%	0.65%
10-50%	6,896	2.19%	3.12%	0.84%	2.31%	3.44%	0.90%	2.55%	3.41%	0.69%
<10%	37,294	1.39%	1.98%	0.33%	1.14%	2.58%	0.42%	1.20%	2.38%	0.20%
% Owner-Occupied DUs in Segment										
50-100%	35,824	1.38%	2.07%	0.35%	1.20%	2.66%	0.47%	1.28%	2.57%	0.25%
10-50%	9,319	2.47%	3.26%	0.75%	1.82%	3.45%	0.77%	1.83%	2.86%	0.48%
<10%	2,945	2.92%	3.77%	0.96%	2.41%	3.80%	0.79%	2.58%	3.58%	0.61%
Combined Median Rent/Housing Value										
1st Quintile	9,184	1.56%	1.67%	0.37%	1.52%	3.16%	0.59%	1.58%	2.87%	0.26%
2nd Quintile	9,846	1.50%	2.32%	0.37%	1.22%	2.43%	0.45%	1.34%	2.80%	0.22%
3rd Quintile	10,105	1.98%	3.28%	0.44%	1.37%	3.23%	0.55%	1.48%	2.59%	0.26%
4th Quintile	9,365	1.74%	2.06%	0.43%	1.60%	3.35%	0.59%	1.56%	2.89%	0.30%
5th Quintile	9,588	1.63%	2.46%	0.65%	1.27%	2.22%	0.54%	1.40%	2.33%	0.49%
Population Density										
Large MSA	16,872	2.11%	2.94%	0.51%	1.79%	3.46%	0.67%	1.99%	3.10%	0.35%
Medium-Small MSA	17,883	1.70%	2.12%	0.50%	1.49%	2.86%	0.55%	1.48%	2.44%	0.33%
Non-MSA, Urban	6,001	1.10%	1.96%	0.32%	0.75%	1.39%	0.36%	0.82%	1.78%	0.25%
Non-MSA, Rural	7,332	1.15%	1.64%	0.30%	0.78%	2.07%	0.23%	0.78%	2.55%	0.19%
Group Quarters										
Group	1,013	3.06%	4.66%	0.79%	0.99%	1.91%	0.75%	1.28%	2.65%	0.62%
Nongroup	47,075	1.65%	2.37%	0.46%	1.40%	2.89%	0.54%	1.47%	2.68%	0.31%
Household Size										
One	5,884	1.00%	1.08%	0.15%	0.59%	1.17%	0.21%	0.68%	0.72%	0.05%
Two	17,232	1.34%	2.23%	0.35%	1.20%	3.22%	0.37%	1.31%	3.26%	0.21%
Three	13,802	2.06%	3.36%	0.85%	1.62%	3.42%	0.99%	1.70%	3.06%	0.58%
Four or more	11,170	2.10%	4.59%	1.03%	1.83%	4.91%	1.49%	1.84%	4.54%	1.00%

[1] Res = Respondent, QDU = questionnaire dwelling unit, NR = nonresponse adjustment, PS = poststratification adjustment.
[2] Weighted extreme value proportion: $100 * \sum_k w_{ek}/\sum_k w_k$, where w_{ek} denotes the weight for extreme values and w_k denotes the weight for both extreme values and non-extreme values.
[3] Outwinsor weight proportion: $100 * \sum_k (w_{ek}-b_k)/\sum_k w_k$, where b_k denotes the winsorized weight.

Appendix F

Evaluation of Calibration Weights: Questionnaire Dwelling Unit–Level Slippage Rates

Table F.1 2002 NSDUH QDU-Level Slippage Rates

Domain	*n*	Initial Total (I)[1]	Final Total (F)[2]	Control from SDU Weights (C)	(I-C)/C%	(F-C)/C %
Total	48,088	109,493,326	109,493,326	109,493,326	0.00	-0.00
Census Region						
Northeast	9,724	20,669,077	20,669,077	20,669,077	0.00	-0.00
South	14,877	39,773,063	39,773,063	39,773,063	0.00	-0.00
Midwest	13,489	25,407,937	25,407,937	25,407,937	0.00	0.00
West	9,998	23,643,248	23,643,248	23,643,248	0.00	0.00
Quarter						
Quarter 1	12,014	27,187,773	27,187,773	27,187,773	0.00	-0.00
Quarter 2	11,692	27,485,169	27,485,169	27,485,169	0.00	0.00
Quarter 3	13,060	27,198,275	27,198,275	27,198,275	0.00	0.00
Quarter 4	11,322	27,622,109	27,622,109	27,622,109	0.00	-0.00
Household Type						
12-17, 18-25, 26+	4,539	4,466,370	4,466,370	4,466,370	0.00	0.00
12-17, 18-25	116	97,383	97,383	97,383	0.00	0.00
12-17, 26+	15,573	13,682,980	13,682,980	13,682,980	0.00	-0.00
18-25, 26+	9,239	11,299,280	11,299,280	11,299,280	0.00	0.00
12-17	72	48,425	48,425	48,425	0.00	0.00
18-25	6,396	6,433,577	6,433,577	6,433,577	0.00	-0.00
26+	12,153	73,465,310	73,465,310	73,465,310	0.00	0.00
Race of Householder						
Hispanic white	5,087	10,229,445	10,229,445	10,229,445	0.00	0.00
Hispanic black	144	461,291	461,290	461,291	-0.00	-0.00
Hispanic other	308	468,411	468,411	468,411	0.00	0.00
Non-Hispanic white	33,802	80,099,867	80,099,867	80,099,867	0.00	0.00
Non-Hispanic black	5,846	12,677,901	12,677,901	12,677,901	0.00	0.00
Non-Hispanic other	2,901	5,556,411	5,556,411	5,556,411	0.00	0.00
% Hispanic in Segment						
50-100%	2,458	6,323,172	6,323,172	6,323,172	0.00	0.00
10-50%	7,856	20,724,410	20,724,410	20,724,410	0.00	-0.00
<10%	37,774	82,445,744	82,445,744	82,445,744	0.00	-0.00
% Black in Segment						
50-100%	3,898	8,535,401	8,535,401	8,535,401	0.00	-0.00
10-50%	6,896	16,987,675	16,987,675	16,987,675	0.00	0.00
<10%	37,294	83,970,250	83,970,250	83,970,250	0.00	-0.00
% Owner-Occupied DUs in Segment						
50-100%	35,824	82,380,762	82,380,762	82,380,762	0.00	-0.00
10-50%	9,319	20,991,291	20,991,291	20,991,291	0.00	0.00
<10%	2,945	6,121,273	6,121,273	6,121,273	0.00	0.00
Combined Median Rent/Housing Value						
1st Quintile	9,184	16,952,021	16,952,021	16,952,021	0.00	0.00
2nd Quintile	9,846	19,953,772	19,953,772	19,953,772	0.00	0.00
3rd Quintile	10,105	23,789,210	23,789,210	23,789,210	0.00	-0.00
4th Quintile	9,365	24,619,692	24,619,692	24,619,692	0.00	0.00
5th Quintile	9,588	24,178,630	24,178,630	24,178,630	0.00	-0.00

Table F.1 2002 NSDUH QDU-Level Slippage Rates (continued)

Domain	*n*	Initial Total (I)[1]	Final Total (F)[2]	Control from SDU Weights (C)	(I-C)/C%	(F-C)/C %
Population Density						
Large MSA	16,872	48,261,484	48,261,484	48,261,484	0.00	-0.00
Medium-Small MSA	17,883	36,601,816	36,601,816	36,601,816	0.00	0.00
Non-MSA, Urban	6,001	10,565,423	10,565,423	10,565,423	0.00	-0.00
Non-MSA, Rural	7,332	14,064,604	14,064,604	14,064,604	0.00	0.00
Group Quarters						
Group	1,013	1,013,804	1,013,804	1,013,804	0.00	0.00
Nongroup	47,075	108,479,522	108,479,522	108,479,522	0.00	-0.00
Household Size						
One	5,884	29,809,582	29,796,890	29,297,697	1.75	1.70
Two	17,232	50,680,040	50,693,122	51,530,620	-1.65	-1.63
Three	13,802	17,028,030	17,037,025	16,818,946	1.24	1.30
Four or more	11,170	11,975,675	11,966,289	11,846,062	1.09	1.01

[1] YR02WT1*...*YR02WT9*DU02WT10*...*DU02WT12 (before QDU poststratification).
[2] YR02WT1*...*YR02WT9*DU02WT10*...*DU02WT13 (after QDU poststratification).

Appendix G

Evaluation of Calibration Weights: Questionnaire Dwelling Unit–Level Weight Summary Statistics

Table G.1 2002 NSDUH Selected QDU-Level Weight Summary Statistics

Domain	*n*	Screener DU-Level Weights (SDUWT: YR02WT1*...*YR02WT9)						Before sel.qdu.ps[1] (SDUWT*DU02WT10)						After sel.qdu.ps[1] (SDUWT*DU02WT11)					
		Min	Q1[2]	Med	Q3[2]	Max	UWE[3]	Min	Q1[2]	Med	Q3[2]	Max	UWE[3]	Min	Q1[2]	Med	Q3[2]	Max	UWE[3]
Total	55,686	14	482	709	1,117	8,832	1.44	14	586	978	2,054	57,237	2.90	13	573	971	2,070	38,523	2.91
Census Region																			
Northeast	11,436	20	360	669	947	8,298	1.44	20	557	844	1,767	27,935	2.92	14	540	843	1,785	38,523	3.00
South	17,121	20	652	921	1,281	8,832	1.32	20	787	1,215	2,436	57,237	2.67	18	773	1,209	2,464	35,120	2.62
Midwest	15,582	15	507	614	780	5,069	1.28	16	565	753	1,758	27,351	2.78	25	553	763	1,737	26,276	2.77
West	11,547	14	262	679	1,447	6,085	1.66	14	399	1,067	2,127	35,038	3.17	13	395	1,055	2,131	34,950	3.28
Quarter																			
Quarter 1	13,703	20	480	711	1,157	7,332	1.44	20	589	981	2,074	36,380	2.90	14	572	966	2,090	31,042	2.93
Quarter 2	13,579	20	499	704	1,115	8,298	1.43	20	589	955	2,006	57,237	3.00	18	592	982	2,070	34,950	2.98
Quarter 3	15,149	14	466	663	1,020	5,684	1.40	14	554	920	1,878	27,935	2.76	13	539	915	1,885	38,523	2.85
Quarter 4	13,255	20	497	754	1,215	8,832	1.47	20	629	1,077	2,274	38,741	2.89	20	608	1,049	2,240	35,120	2.86
Household Type																			
12-17, 18-25, 26+	4,959	20	523	752	1,192	7,462	1.46	20	523	752	1,192	7,462	1.46	14	502	739	1,206	10,567	1.53
12-17, 18-25	131	71	425	710	989	2,387	1.35	71	425	710	989	2,388	1.35	66	376	682	988	3,703	1.44
12-17, 26+	17,089	14	440	665	1,048	7,910	1.47	14	444	672	1,061	7,910	1.47	13	422	666	1,068	11,074	1.53
18-25, 26+	10,878	15	548	773	1,217	8,298	1.41	16	593	863	1,375	9,340	1.45	25	565	846	1,383	14,808	1.52
12-17	78	36	216	604	878	2,001	1.48	36	216	610	927	2,021	1.46	42	214	515	900	2,119	1.60
18-25	7,062	24	486	700	1,086	8,832	1.43	26	515	772	1,231	8,902	1.45	23	508	784	1,194	10,058	1.48
26+	15,489	39	480	704	1,115	7,332	1.42	120	2,148	3,709	6,252	57,237	1.66	115	2,131	3,728	6,297	38,523	1.66
Race of Householder																			
Hispanic white	5,798	20	573	874	1,352	7,332	1.34	20	631	1,103	1,815	36,380	2.50	15	641	1,095	1,834	27,533	2.54
Hispanic black	163	43	686	1,280	1,815	8,298	1.81	43	839	1,588	3,316	26,273	2.65	60	893	1,408	3,735	31,056	2.63
Hispanic other	351	14	168	430	1,076	7,024	2.28	14	221	560	1,433	24,367	4.24	13	232	567	1,391	15,253	3.53
Non-Hispanic white	39,367	22	476	682	1,061	5,498	1.41	22	583	967	2,182	29,762	2.88	14	571	960	2,182	34,950	2.91
Non-Hispanic black	6,616	55	578	812	1,165	7,462	1.36	55	656	989	1,786	57,237	3.03	27	641	998	1,853	34,346	2.97
Non-Hispanic other	3,391	20	231	570	1,182	8,832	1.88	20	312	822	1,894	38,741	3.31	14	313	794	1,845	38,523	3.29
% Hispanic in Segment																			
50-100%	2,817	41	707	1,095	1,480	7,910	1.27	41	812	1,356	2,070	27,575	2.41	36	841	1,415	2,200	21,762	2.43
10-50%	9,109	22	587	937	1,436	8,298	1.37	22	702	1,264	2,308	35,038	2.66	14	688	1,259	2,352	34,950	2.73
<10%	43,760	14	451	666	1,009	8,832	1.44	14	559	903	1,965	57,237	2.99	13	547	901	1,992	38,523	2.99

Table G.1 2002 NSDUH Selected QDU-Level Weight Summary Statistics (continued)

Domain	*n*	Screener DU-Level Weights (SDUWT: YR02WT1*...*YR02WT9)						Before sel.qdu.ps[1] (SDUWT*DU02WT10)						After sel.qdu.ps[1] (SDUWT*DU02WT11)					
		Min	Q1[2]	Med	Q3[2]	Max	UWE[3]	Min	Q1[2]	Med	Q3[2]	Max	UWE[3]	Min	Q1[2]	Med	Q3[2]	Max	UWE[3]
% Black in Segment																			
50-100%	4,429	20	545	773	1,135	8,832	1.42	20	624	958	1,836	57,237	3.25	18	614	978	1,871	34,950	3.05
10-50%	7,935	20	594	853	1,292	7,462	1.38	20	690	1,139	2,214	38,741	2.76	14	673	1,128	2,196	35,120	2.76
<10%	43,322	14	451	680	1,078	7,910	1.45	14	562	951	2,044	35,038	2.90	13	552	941	2,059	38,523	2.93
%Owner-Occupied DUs in Segment																			
50-100%	41,619	14	478	691	1,088	8,832	1.43	14	577	967	2,107	57,237	2.91	13	564	955	2,117	38,523	2.93
10-50%	10,720	20	494	759	1,186	8,298	1.46	20	613	1,010	1,947	36,380	2.88	14	608	1,021	2,007	31,042	2.86
<10%	3,347	20	525	795	1,230	6,085	1.44	20	616	989	1,849	27,575	2.85	23	599	1,014	1,838	22,075	2.81
Combined Median Rent/Housing Value																			
1st Quintile	10,526	15	358	593	812	4,170	1.46	16	504	775	1,717	27,807	2.93	23	483	778	1,736	27,883	2.93
2nd Quintile	11,363	14	322	617	967	8,832	1.60	14	478	827	1,875	38,741	3.20	13	464	811	1,849	35,120	3.16
3rd Quintile	11,690	22	537	759	1,105	7,462	1.37	22	637	1,012	2,147	36,380	2.84	14	630	1,009	2,158	31,151	2.83
4th Quintile	10,892	22	600	883	1,289	5,647	1.34	22	703	1,174	2,264	57,237	2.72	20	698	1,171	2,335	38,523	2.74
5th Quintile	11,215	20	544	774	1,259	8,298	1.41	20	632	1,136	2,205	27,623	2.76	18	625	1,127	2,260	32,355	2.83
Population Density																			
Large MSA	19,841	20	636	893	1,349	7,910	1.30	20	740	1,270	2,621	29,762	2.54	18	729	1,271	2,639	38,523	2.59
Medium-Small MSA	20,648	14	404	656	999	8,832	1.46	14	535	897	1,849	57,237	2.95	13	527	895	1,871	34,346	2.90
Non-MSA, Urban	6,771	31	247	582	892	5,992	1.56	31	378	766	1,510	36,380	3.55	20	364	748	1,477	27,533	3.50
Non-MSA, Rural	8,426	15	260	582	897	4,183	1.58	16	417	771	1,683	38,741	3.28	16	410	775	1,714	35,120	3.30
Group Quarters																			
Group	1,046	24	345	638	1,041	2,876	1.50	26	366	698	1,124	13,292	2.59	23	389	731	1,111	13,478	2.45
Nongroup	54,640	14	485	711	1,119	8,832	1.44	14	588	986	2,081	57,237	2.89	13	576	979	2,102	38,523	2.90
Household Size																			
One	7,101	24	452	664	1,046	4,528	1.40	26	811	2,332	6,269	57,237	2.27	23	809	2,280	6,138	38,523	2.26
Two	20,520	25	491	708	1,099	8,832	1.40	25	693	1,368	3,261	26,785	2.23	16	683	1,351	3,273	34,950	2.28
Three	15,742	15	485	713	1,116	8,298	1.43	16	516	784	1,284	24,367	2.03	14	503	778	1,297	17,372	2.13
Four or more	12,323	14	486	734	1,199	7,910	1.51	14	495	750	1,246	11,751	1.66	13	471	745	1,256	11,988	1.76

[1] Sel = selected, QDU = questionnaire dwelling unit, PS = poststratification.
[2] Q1 and Q3 refer to the first and third quartile of the weight distribution.
[3] Unequal weighting effect defined as $1+[(n-1)/n]*CV^2$, where CV=coefficient of variation of weights.

Table G.2 2002 NSDUH Respondent QDU-Level Weight Summary Statistics

Domain	*n*	Before res.qdu.nr[1] (SDUWT*DU02WT10*...*DU02WT11)						After res.qdu.nr[1] (SDUWT*DU02WT10*...*DU02WT12)						Final Weight: After res.qdu.ps[1] (SDUWT*DU02WT10*...*DU02WT13)					
		Min	Q1[2]	Med	Q3[2]	Max	UWE[3]	Min	Q1[2]	Med	Q3[2]	Max	UWE[3]	Min	Q1[2]	Med	Q3[2]	Max	UWE[3]
Total	48,088	13	560	935	1,903	35,120	2.93	13	620	1,054	2,201	52,370	3.36	12	619	1,055	2,201	53,141	3.35
Census region																			
Northeast	9,724	14	518	802	1,583	25,521	3.00	14	568	906	1,870	41,057	3.63	12	564	908	1,891	40,953	3.64
South	14,877	18	755	1,171	2,225	35,120	2.64	18	837	1,301	2,560	44,217	2.98	17	837	1,303	2,568	44,633	2.97
Midwest	13,489	25	547	741	1,559	26,276	2.83	25	607	848	1,820	31,590	3.04	23	607	853	1,821	31,575	3.04
West	9,998	13	379	1,004	2,020	32,266	3.26	13	424	1,134	2,299	52,370	3.84	12	424	1,133	2,307	53,141	3.84
Quarter																			
Quarter 1	12,014	14	562	935	1,945	27,533	2.94	14	614	1,041	2,199	42,256	3.29	12	614	1,042	2,191	43,090	3.30
Quarter 2	11,692	18	576	938	1,909	34,346	3.01	18	643	1,063	2,208	44,207	3.45	17	642	1,063	2,212	48,756	3.45
Quarter 3	13,060	13	528	875	1,711	32,266	2.87	13	581	986	1,976	44,217	3.34	12	579	986	1,977	44,633	3.35
Quarter 4	11,322	20	592	1,004	2,072	35,120	2.85	20	665	1,140	2,451	52,370	3.29	17	666	1,141	2,451	53,141	3.28
Household Type																			
12-17, 18-25, 26+	4,539	14	498	731	1,200	10,567	1.53	14	538	796	1,313	10,033	1.56	12	532	800	1,319	8,784	1.54
12-17, 18-25	116	66	375	662	949	2,120	1.34	67	420	723	1,192	2,387	1.40	65	442	722	1,186	2,384	1.40
12-17, 26+	15,573	13	419	665	1,068	11,074	1.53	13	460	731	1,171	10,539	1.54	12	458	730	1,173	8,440	1.52
18-25, 26+	9,239	25	560	841	1,377	14,808	1.53	31	639	986	1,607	22,351	1.60	26	639	990	1,608	14,076	1.55
12-17	72	42	210	515	936	2,119	1.61	42	218	568	1,034	2,354	1.61	41	194	567	1,051	2,343	1.66
18-25	6,396	23	505	785	1,205	10,058	1.48	26	551	867	1,318	9,530	1.47	25	549	866	1,319	6,652	1.47
26+	12,153	121	2,088	3,653	6,146	35,120	1.66	123	2,556	4,592	7,986	52,370	1.76	116	2,552	4,597	7,995	53,141	1.76
Race of Householder																			
Hispanic white	5,087	15	631	1,087	1,781	27,533	2.53	15	695	1,214	2,037	29,947	2.77	14	696	1,222	2,041	29,707	2.77
Hispanic black	144	60	847	1,328	3,276	31,056	2.81	60	979	1,626	3,346	44,207	3.23	60	974	1,538	3,143	45,894	3.43
Hispanic other	308	13	224	539	1,390	15,253	3.67	13	239	661	1,608	19,209	3.88	12	242	647	1,675	18,181	3.85
Non-Hispanic white	33,802	14	555	920	1,980	29,897	2.93	14	618	1,038	2,319	52,370	3.37	13	619	1,039	2,318	53,141	3.36
Non-Hispanic black	5,846	27	634	972	1,772	34,346	2.99	36	687	1,074	1,970	42,296	3.35	31	685	1,075	1,982	43,059	3.35
Non-Hispanic other	2,901	14	296	738	1,717	35,120	3.25	14	327	856	1,992	42,256	3.83	12	324	868	2,024	43,090	3.88
% Hispanic in Segment																			
50-100%	2,458	36	828	1,380	2,084	21,373	2.43	36	928	1,531	2,402	39,693	2.77	36	927	1,535	2,398	39,900	2.76
10-50%	7,856	14	669	1,222	2,150	32,266	2.73	14	743	1,379	2,530	48,233	3.17	12	742	1,379	2,530	48,756	3.18
<10%	37,774	13	535	864	1,796	35,120	3.01	13	591	974	2,077	52,370	3.44	12	592	977	2,081	53,141	3.43

Table G.2 2002 NSDUH Respondent QDU-Level Weight Summary Statistics (continued)

Domain	*n*	Before res.qdu.nr[1] (SDUWT*DU02WT10*...*DU02WT11)						After res.qdu.nr[1] (SDUWT*DU02WT10*...*DU02WT12)						Final Weight: After res.qdu.ps[1] (SDUWT*DU02WT10*...*DU02WT13)					
		Min	Q1[2]	Med	Q3[2]	Max	UWE[3]	Min	Q1[2]	Med	Q3[2]	Max	UWE[3]	Min	Q1[2]	Med	Q3[2]	Max	UWE[3]
% Black in Segment																			
50-100%	3,898	18	608	956	1,785	34,346	2.98	18	654	1,066	1,982	43,351	3.38	17	651	1,068	1,997	48,756	3.38
10-50%	6,896	14	658	1,096	2,063	35,120	2.81	14	733	1,236	2,377	42,296	3.13	12	733	1,235	2,380	43,059	3.13
<10%	37,294	13	539	903	1,879	31,056	2.95	13	597	1,017	2,182	52,370	3.40	12	598	1,019	2,181	53,141	3.39
% Owner-Occupied DUs in Segment																			
50-100%	35,824	13	550	918	1,930	35,120	2.95	13	610	1,029	2,253	52,370	3.37	12	611	1,032	2,255	53,141	3.37
10-50%	9,319	14	596	992	1,873	28,184	2.87	14	651	1,120	2,141	42,256	3.29	12	648	1,122	2,135	43,090	3.30
<10%	2,945	23	585	994	1,775	22,075	2.81	23	633	1,102	1,996	39,693	3.29	22	630	1,099	2,002	39,900	3.29
Combined Median Rent/Housing Value																			
1st Quintile	9,184	23	464	750	1,564	27,883	2.99	26	517	844	1,816	42,296	3.36	25	516	845	1,821	43,059	3.36
2nd Quintile	9,846	13	452	784	1,733	35,120	3.19	13	500	882	2,011	35,739	3.54	12	500	886	2,004	35,915	3.54
3rd Quintile	10,105	14	618	968	1,958	31,056	2.82	14	677	1,095	2,269	52,370	3.26	12	678	1,098	2,267	53,141	3.25
4th Quintile	9,365	20	684	1,127	2,133	34,346	2.71	20	765	1,264	2,483	43,351	3.18	17	765	1,263	2,490	48,756	3.17
5th Quintile	9,588	18	610	1,083	2,048	26,537	2.87	18	678	1,212	2,403	48,233	3.28	17	677	1,212	2,393	48,330	3.29
Population Density																			
Large MSA	16,872	18	713	1,220	2,390	32,266	2.59	18	804	1,375	2,847	52,370	2.99	17	804	1,375	2,854	53,141	2.99
Medium-Small MSA	17,883	13	512	864	1,711	34,346	2.89	13	575	978	1,980	39,693	3.31	12	576	982	1,986	39,900	3.30
Non-MSA, Urban	6,001	20	356	721	1,412	27,533	3.59	20	382	801	1,560	31,590	3.88	17	381	805	1,568	31,575	3.88
Non-MSA, Rural	7,332	16	387	743	1,609	35,120	3.37	17	432	838	1,850	44,217	3.84	15	434	839	1,854	44,633	3.83
Group Quarters																			
Group	1,013	23	391	725	1,120	13,478	2.46	26	392	757	1,134	15,308	2.50	25	399	755	1,138	15,339	2.49
Nongroup	47,075	13	562	942	1,929	35,120	2.92	13	623	1,065	2,241	52,370	3.34	12	622	1,066	2,239	53,141	3.33
Household Size																			
One	5,884	23	772	2,021	5,712	35,120	2.32	26	876	2,399	7,212	52,370	2.50	25	876	2,394	7,205	53,141	2.50
Two	17,232	16	659	1,255	2,987	23,819	2.31	25	743	1,441	3,710	32,508	2.57	23	745	1,441	3,702	33,097	2.57
Three	13,802	14	497	767	1,276	15,253	2.07	14	549	868	1,428	29,031	2.47	12	551	871	1,428	30,071	2.46
Four or more	11,170	13	463	735	1,238	11,988	1.73	13	504	812	1,370	22,166	1.95	12	502	814	1,372	22,898	1.93

[1] Res = respondent, QDU = questionnaire dwelling unit, PS = poststratification adjustment, NR = nonresponse adjustment

[2] Q1 and Q3 refer to the first and third quartile of the weight distribution.

[3] Unequal weighting effect defined as $1+[(n-1)/n]*CV^2$, where CV=coefficient of variation of weights.

Appendix H

GEM Modeling Summary for the Pair Weights

Appendix H

GEM Modeling Summary for the Pair Weights

Introduction

This appendix summarizes each model group throughout all stages of weight calibration modeling. Unlike much of the other information presented in this report, this section provides a model-specific overview of weight calibration, as opposed to a domain-specific one.

For 2002, modeling involved taking two model groups through four adjustment steps: 1) selected pair poststratification, 2) pair nonresponse adjustment, 3) responding pair poststratification, and 4) responding pair extreme value adjustment.

Model-specific summary statistics are shown in Tables H1a, H1b to H2a, and H2b. Included in these tables, for each stage of modeling, are: the number of factor effects included in the final model; the high, low, and nonextreme weight bounds set to provide the upper and lower limits for the generalized exponential model (GEM) macro; the weighted, unweighted and winsorized weight proportions; the unequal weighting effect (UWE); and weight distributions. The UWE provides an approximate partial measure of variance and provides a summary of how much impact a particular stage of modeling has on the distribution of the new product of weights. At each stage in the modeling, these summary statistics were calculated and utilized to help evaluate the quality of the weight component under the model chosen.

Occurrences of small sample sizes and exact linear combinations in the realized data lead to situations whereby modeling inclusion of all originally proposed levels of covariates in the model is not possible. The text and exhibits in Sections H1 and H2 summarize the decisions made with regard to final covariates included in each model. For a list of the proposed initial covariates considered at each stage of modeling, see Exhibit H.1; for the list of realized final model covariates, see Exhibits H1.1 to H2.4. For guidelines on interpreting these exhibits, see Appendix C.

H.1 Final Model Explanatory Variables

For brevity, numeric abbreviations for factor levels are established in Exhibit 4.2 (included here as Exhibit H for easy reference) in Chapter 4. There, a complete list is provided of all variables and associated levels used at any stage of modeling. Note that not all factors or levels are present in all stages of modeling, and the initial set of variables is the same across model groups but may change for an adjustment step of modeling. The initial candidates are found in any of the proposed variables columns for a particular stage of weight adjustment.

Exhibit H.1 Definitions of Levels for Pair-Level Calibration Modeling Variables

Group Quarter Indicator

1: College Dorm, 2: Other Group Quarter, 3: Nongroup Quarter [1]

Household Size

2: DU with 2 persons, 3: DU with 3 persons, 4: DU with >=4 persons

Pair Age (15 levels)

1: 12-17 & 12-17 [1], 2: 12-17 & 18-25, 3: 12-17 & 26-34, 4: 12-17 & 35-49, 5: 12-17 & 50+,
6: 18-25 & 18-25, 7: 18-25 & 26-34, 8: 18-25 & 35-49, 9: 18-25 & 50+, 10: 26-34 & 26-34,
11: 26-34 & 35-49, 12: 26-34 & 50+, 13: 35-49 & 35-49, 14: 35-49 & 50+, 15: 50+ & 50+

Pair Age (6 levels)

1: 12-17 & 12-17 [1], 2: 12-17 & 18-25, 3: 12-17 & 26+, 4: 18-25 & 18-25, 5: 18-25 & 26+, 6: 26+ & 26+.

Pair Age (3 levels)

1: 12-17 & 12-17 [1], 2: 12-17 & 18+, 3: 18+ & 18+

Pair Gender

1: Male & Female [1], 2: Female & Female, 3: Male & Male

Pair Race (10 levels)

1: white & white [1], 2: white & black, 3: white & Hispanic, 4: white & other, 5: black & black,
6: black & Hispanic, 7: black & other, 8: Hispanic & Hispanic, 9: Hispanic & other, 10: other & other.

Pair Race (5 levels)

1: Mixed race pair, 2: Hispanic pair, 3: black pair, 4: white pair[1], 5: other pair.

Pair Race (4 levels)

1: Mixed race pair or other & other, 2: Hispanic pair, 3: black pair, 4 white pair[1]

Percentage of Owner-Occupied Dwelling Units in Segment (% Owner-Occupied)

1: 50%-100% [1], 2: 10%->50%, 3: 0->10%

Percentage of Segments That Are Black (% black)

1: 50%-100%, 2: 10%->50%, 3: 0->10% [1]

Percentage of Segments That Are Hispanic (% Hispanic)

1: 50%-100%, 2: 10%->50%, 3: 0->10% [1]

Segment-Combined Median Rent and Housing Value (Rent/Housing)[2]

1: First Quintile, 2: Second Quintile, 3: Third Quintile, 4: Fourth Quintile, 5: Fifth Quintile[1]

Population Density

1: MSA 1,000,000 or more, 2: MSA less than 1,000,000, 3: Non-MSA urban, 4: Non-MSA rural [1]

Quarter

1: Quarter 1, 2: Quarter 2, 3: Quarter 3, 4: Quarter 4 [1]

Race of Householder

1: Hispanic white [1], 2: Hispanic black, 3: Hispanic others, 4: Non-Hispanic white,
5: Non-Hispanic black, 6: Non-Hispanic others,

State / Region

Model Group 1**:** 1: Connecticut, Maine, Massachusetts, New Hampshire, New Jersey, Rhode Island, Vermont,
2: Alabama, Arkansas, Delaware, District of Columbia, Georgia, Kentucky, Louisiana, Maryland, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee, Virginia, West Virginia [1], 3: New York, 4: Pennsylvania, 5: Florida, 6: Texas,

Model Group 2: 1: Indiana, Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota, South Dakota,| Wisconsin, 2: Alaska, Arizona, Colorado, Idaho, Hawaii, Montana, Nevada, New Mexico, Oregon, Utah, Washington, Wyoming [1], 3: Michigan, 4: Illinois, 5: Ohio, 6: California

Exhibit H.1 Definitions of Levels for Pair-Level Calibration Modeling Variables (continued)

States [3]

Model Group 1: 1: Alabama, 2: Arkansas, 3: Connecticut, 4: Delaware, 5: District of Columbia, 6: Florida, 7: Georgia, 8: Kentucky, 9: Louisiana, 10: Maine, 11: Maryland, [1] 12: Massachusetts, 13: Mississippi, 14: New Hampshire, 15: New Jersey, 16: New York, 17: North Carolina, 18: Oklahoma, 19: Pennsylvania, 20: Rhode Island, 21: South Carolina, 22: Tennessee, 23: Texas, 24: Vermont, 25: Virginia, 26: West Virginia

Model Group 2: 1: Alaska, 2: Arizona, [1] 3: California, 4: Colorado, 5: Idaho, 6: Illinois, 7: Indiana, 8: Iowa, 9: Hawaii, 10: Kansas, 11: Michigan, 12: Minnesota, 13: Missouri, 14: Montana, 15: Nebraska, 16: Nevada, 17: New Mexico, 18: North Dakota, 19: Ohio, 20: Oregon, 21: South Dakota, 22: Utah, 23: Washington, 24: Wisconsin, 25: Wyoming

Pair Relationship Associated with Multiplicity

1: Parent-child (12-14)*
2: Parent-child (12-17)*
3: Parent-child (12-10)*
4: Parent*-child (12-14)
5: Parent*-child (12-17)
6: Parent*-child (12-20)
7: Sibling (12-14)-sibling (15-17)
8: Sibling (12-17)-sibling (18-25)
9: Spouse-spouse
10: Spouse-spouse with kids

[1] The reference level for this variable. This is the level against which effects of other factor levels are measured.

[2] Segment-combined Median Rent and Housing Value is a composite measure based on rent, housing value, and percent Owner-Occupied.

[3] The States or district assigned to a particular model is based on combined Census regions.

Exhibit H.2 Covariates for 2002 NSDUH Pair Weights

Variables	Level	Proposed
One-Factor Effects		
Intercept	1	1
State	Model Specific	
Quarter	4	3
Population Density	3	2
Group Quarter	3	2
Household size	3	2
Pair Age	15	14
Pair Sex	4	2
Pair Race	10	9
Race of Householder	6	5
Rent/housing	5	4
Segment % black	3	2
Segment % Hispanic	3	2
% Owner-Occupied	3	2
Pair Relationship	Model Specific	
Two-Factor Effects		
Pair Race (5 levels) x Pair Age (6 levels)	5 x 6	20
Pair Race (5 levels) x Pair Sex	5 x 3	8
Pair Sex x Pair Age (6 levels)	3 x 6	10
State/Region x Pair Race (5 levels)	Model Specific	
State/Region x Pair Age (6 levels)	Model Specific	
State/Region x Pair Sex	Model Specific	
Rent/housing x % black	5 x 3	8
Rent/housing x % Hispanic	5 x 3	8
Rent/housing x % Owner-Occupied	5 x 3	8
% Owner-Occupied x % black	3 x 3	4
% Owner-Occupied x % Hispanic	3 x 3	4
Three-Factor Effects		
Pair Race (4 levels) x Pair Sex x Pair Age (3 levels)	4 x 3 x 3	12

Appendix H1

Model Group 1: Northeast and South

Alabama, Arkansas, Connecticut, Delaware, District of Columbia, Florida, Georgia, Kentucky, Louisiana, Maine, Massachusetts, Maryland, Mississippi, New Hampshire, New Jersey, New York, North Carolina, Oklahoma, Pennsylvania, Rhode Island, South Carolina, Tennessee, Texas, Vermont, Virginia, West Virginia

Table H1a 2002 Pair Weight GEM Modeling Summary (Model Group 1: Northeast and South)

Modeling Step[1]	Extreme Weight Proportions			UWE[2]	# XVAR[3]	Bounds[4]	
	Unweighted	Weighted	Winsorized			Nominal	Realized
sel.pr.ps	5.96%	8.98%	20.20%	45.317	213	(0.4, 2.8)	(0.40, 2.80)
	3.30%	13.31%	2.35%	12.189	154	(0.2, 3.5)	(0.20, 3.47)
						(0.9, 1.1)	(0.90, 1.10)
res.pr.nr	3.43%	10.90%	2.10%	16.307	213	(1.0, 5.0)	(1.00, 5.00)
	2.54%	15.23%	3.87%	15.238	213	(1.0, 5.5)	(1.00, 5.50)
						n/a	n/a
res.pr.ps	2.56%	16.50%	5.43%	15.238	223	(0.38, 2.95)	(0.40, 2.86)
	2.49%	12.48%	2.66%	11.645	164	(0.28, 2.95)	(0.30, 2.59)
						(0.81, 1.1)	(0.81, 0.81)
res.pr.ev	2.49%	12.48%	2.66%	11.645	223	(0.4, 2.0)	(0.87, 1.82)
	1.24%	5.96%	0.77%	11.630	164	(0.6, 2.0)	(0.66, 1.75)
						(0.9, 1.1)	(0.93, 0.93)

[1] For a key to modeling abbreviations, see Chapter 7, Exhibit 7.2.

[2] Unequal weighting effect defined as $1+[(n-1)/n]*CV^2$, where CV=coefficient of variation of weights.

[3] Number of proposed covariates on top line, and number finalized after modeling.

[4] Nominal bounds are used in defining maximum/minimum values for the GEM adjustment factors. The realized bound is the actual adjustment produced by the modeling. The first set of bounds listed is for high extreme values, the second for nonextreme, and the third for low extreme values.

Table H1b 2002 Distribution of Weight Adjustment Factors and Weight Products (Model Group 1: Midwest and West)

	SDU wt	pair selection prob		sel.pr.ps [1]		res.pr.nr [1]		res.pr.ps [1]		res.pr.ev [1]	
	1-9	pairwt10	1-10	pairwt11	1-11	pairwt12	1-12	pairwt13	1-13	pairwt14	1-14
Minimum	20	1.02	25	0.07	6	0.41	6	0.18	2	0.51	2
1%	73	1.02	138	0.21	110	0.85	10	0.34	81	0.77	74
5%	143	1.14	347	0.31	261	1.00	270	0.57	230	0.85	213
10%	221	1.20	655	0.45	430	1.01	445	0.70	376	0.88	352
25%	556	1.38	1,104	0.67	999	1.03	1,037	0.86	947	0.94	910
Median	839	4.49	2,767	1.00	2,788	1.11	2,960	0.98	2,899	0.99	2,866
75%	1,191	10.51	8,351	1.37	7,926	1.26	8,494	1.11	8,538	1.02	8,541
90%	1,620	17.77	17,895	1.76	20,156	1.52	22,818	1.26	22,835	1.06	23,085
95%	1,917	29.60	27,873	2.01	33,883	1.82	43,234	1.39	43,242	1.09	43,363
99%	2,683	50.55	59,580	2.61	76,999	3.07	120,389	1.77	114,672	1.17	117,503
Maximum	8,832	5,691.21	5,413,350	3.47	1,575,010	5.50	2,072,830	2.59	1,506,420	1.75	1,415,150
n	12,463	-	12,463	-	12,463	-	10,005	-	10,005	-	10,005
mean	913	8.73	7,963	1.06	8,399	1.23	10,463	0.99	10,463	0.98	10,463
Max/mean	9.68	-	680	-	188	-	198	-	144	-	135

[1] For a key to modeling abbreviations, see Chapter 7, Exhibit 7.2.

Model Group 1 Overview

Selected Pair–Level Poststratification

All 76 proposed main effects were included in the model.

No two-factor effects involving State/region were kept in the model due to problems with convergence. Out of 125 two-factor effects, 70 were retained in the model.

The three-factor interaction of pair race by pair sex by pair age was simplified by collapsing. Here the mixed pair race category was collapsed with the black pair race category for all 12 proposed variables. As a result, out of 12 three-factor effects, 8 were kept in the model.

Respondent Pair–Level Nonresponse

All proposed factors were retained in the final model.

Respondent Pair–Level Poststratification

All 86 proposed main effects were included in the model.

No two-factor effects involving State/region were kept in the model due to problems with convergence. Out of 125 two-factor effects, 70 were retained in the model.

The three-factor interaction of pair race by pair sex by pair age was simplified by collapsing. Here the mixed pair race category was collapsed with the black pair race category for all 12 proposed variables. As a result, out of 12 three-factor effects, 8 were kept in the model.

Respondent Pair–Level Extreme Value Adjustment

This step used exactly the same variables as in the respondent pair–level poststratification step.

Exhibit H1.1 Covariates for 2002 NSDUH Pair Weights (sel.pr.ps) Model Group 1: Northeast and South

Variables	Level	Proposed	Final	Comments
One-Factor Effects		**76**	**76**	
Intercept	1	1	1	All levels present.
State	26	25	24	All levels present.
Quarter	4	3	3	All levels present.
Population Density	4	3	3	All levels present.
Group Quarter	3	2	2	All levels present.
Household size	3	2	2	All levels present.
Pair Age	15	14	14	All levels present.
Pair Sex	3	2	2	All levels present.
Pair Race	10	9	9	All levels present.
Race of Householder	6	5	5	All levels present.
Rent/housing	5	4	4	All levels present.
Segment % black	3	2	2	All levels present.
Segment % Hispanic	3	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
Two-Factor Effects		**125**	**70**	
Pair Race (5 levels) x Pair Age (6 levels)	5 x 6	20	20	All levels present.
Pair Race (5 levels) x Pair Sex	5 x 3	8	8	All levels present.
Pair Sex x Pair Age (6 levels)	3 x 6	10	10	All levels present.
State/Region x Pair Race (5 levels)	6 x 5	20	0	Drop all; conv.
State/Region x Pair Age (6 levels)	6 x 6	25	0	Drop all; conv.
State/Region x Pair Sex	6 x 3	10	0	Drop all; conv.
Rent/housing x % black	5 x 3	8	8	All levels present.
Rent/housing x % Hispanic	5 x 3	8	8	All levels present.
Rent/housing x % Owner-Occupied	5 x 3	8	8	All levels present.
% Owner-Occupied x % black	3 x 3	4	4	All levels present.
% Owner-Occupied x % Hispanic	3 x 3	4	4	All levels present.
Three-Factor Effects		**12**	**8**	
Pair Race (4 levels) x Pair Sex x Pair Age (3 levels)	4 x 3 x 3	12	8	Coll. (1,1,1) & (3,1,1); conv. Repeat for all levels of age.
Total		**212**	**154**	

Exhibit H1.2 Covariates for 2002 NSDUH Pair Weights (res.pr.nr) Model Group 1: Northeast and South

Variables	Level	Proposed	Final	Comments
One-Factor Effects		**76**	**76**	
Intercept	1	1	1	All levels present.
State	26	25	25	All levels present.
Quarter	4	3	3	All levels present.
Population Density	4	3	3	All levels present.
Group Quarter	3	2	2	All levels present.
Household size	3	2	2	All levels present.
Pair Age	15	14	14	All levels present.
Pair Sex	3	2	2	All levels present.
Pair Race	10	9	9	All levels present.
Race of Householder	6	5	5	All levels present.
Rent/housing	5	4	4	All levels present.
Segment % black	3	2	2	All levels present.
Segment % Hispanic	3	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
Two-Factor Effects		**125**	**125**	
Pair Race (5 levels) x Pair Age (6 levels)	5 x 6	20	20	All levels present.
Pair Race (5 levels) x Pair Sex	5 x 3	8	8	All levels present.
Pair Sex x Pair Age (6 levels)	3 x 6	10	10	All levels present.
State/Region x Pair Race (5 levels)	6 x 5	20	20	All levels present.
State/Region x Pair Age (6 levels)	6 x 6	25	25	All levels present.
State/Region x Pair Sex	6 x 3	10	10	All levels present.
Rent/housing x % black	5 x 3	8	8	All levels present.
Rent/housing x % Hispanic	5 x 3	8	8	All levels present.
Rent/housing x % Owner-Occupied	5 x 3	8	8	All levels present.
% Owner-Occupied x % black	3 x 3	4	4	All levels present.
% Owner-Occupied x % Hispanic	3 x 3	4	4	All levels present.
Three-Factor Effects		**12**	**12**	
Pair Race (4 levels) x Pair Sex x Pair Age (3 levels)	4 x 3 x 3	12	12	All levels present.
Total		**213**	**213**	

Exhibit H1.3 Covariates for 2002 NSDUH Pair Weights (res.pr.ps) Model Group 1: Northeast and South

Variables	Level	Proposed	Final	Comments
One-Factor Effects		**86**	**86**	
Intercept	1	1	1	All levels present.
State	26	24	25	All levels present.
Quarter	4	3	3	All levels present.
Population Density	4	3	3	All levels present.
Group Quarter	3	2	2	All levels present.
Household size	3	2	2	All levels present.
Pair Age	15	14	14	All levels present.
Pair Sex	3	2	2	All levels present.
Pair Race	10	9	9	All levels present.
Race of Householder	6	5	5	All levels present.
Rent/housing	5	4	4	All levels present.
Segment % black	3	2	2	All levels present.
Segment % Hispanic	3	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
Pair Relationship	10	10	10	All levels present.
Two-Factor Effects		**125**	**70**	
Pair Race (5 levels) x Pair Age (6 levels)	5 x 6	20	20	All levels present.
Pair Race (5 levels) x Pair Sex	5 x 3	8	8	All levels present.
Pair Sex x Pair Age (6 levels)	3 x 6	10	10	All levels present.
State/Region x Pair Race (5 levels)	6 x 5	20	0	Drop all; conv.
State/Region x Pair Age (6 levels)	6 x 6	25	0	Drop all; conv.
State/Region x Pair Sex	6 x 3	10	0	Drop all; conv.
Rent/housing x % black	5 x 3	8	8	All levels present.
Rent/housing x % Hispanic	5 x 3	8	8	All levels present.
Rent/housing x % Owner-Occupied	5 x 3	8	8	All levels present.
% Owner-Occupied x % black	3 x 3	4	4	All levels present.
% Owner-Occupied x % Hispanic	3 x 3	4	4	All levels present.
Three-Factor Effects		**12**	**8**	
Pair Race (4 levels) x Pair Sex x Pair Age (3 levels)	4 x 3 x 3	12	8	Coll. (1,1,1) & (3,1,1); conv. Repeat for all levels of age.
Total		**223**	**164**	

Exhibit H1.4 Covariates for 2002 NSDUH Pair Weights (res.pr.ev) Model Group 1: Northeast and South

This step used the same variables as the respondent pair–level poststratification step in Exhibit H1.3.

Appendix H2

Model Group 2: Midwest and West

Alaska, Arizona, California, Colorado, Hawaii, Idaho, Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Missouri, Montana, Nebraska, Nevada, New Mexico, North Dakota, Ohio, Oregon, South Dakota, Utah, Washington, Wisconsin, Wyoming

Table H2a 2002 Pair Weight GEM Modeling Summary (Model Group 2: Midwest and West)

Modeling Step[1]	Extreme Weight Proportions			UWE[2]	# XVAR[3]	Bounds[4]	
	Unweighted	Weighted	Winsorized			Nominal	Realized
sel.pr.ps	5.86%	4.75%	16.20%	18.663	212	(0.8, 1.7)	(0.80, 1.70)
	2.28%	4.02%	0.60%	12.092	145	(0.3, 3.0)	(0.30, 3.00)
						(0.9, 1.1)	(0.90, 1.10)
res.pr.nr	2.46%	5.02%	0.75%	13.546	212	(1.0, 3.0)	(1.00, 3.00)
	2.14%	10.19%	2.42%	13.844	212	(1.0, 5.0)	(1.00, 4.99)
						n/a	n/a
res.pr.ps	2.24%	13.30%	2.68%	13.844	222	(0.69, 1.67)	(0.71, 1.65)
	1.63%	9.48%	1.12%	14.121	155	(0.44, 1.77)	(0.45, 1.72)
						n/a	n/a
res.pr.ev	1.63%	9.48%	1.12%	14.121	222	(0.5, 1.2)	(0.56, 1.20)
	0.29%	1.31%	0.10%	13.556	155	(0.8, 1.6)	(0.81, 1.59)
						n/a	n/a

[1] For a key to modeling abbreviations, see Chapter 7, Exhibit 7.2.

[2] Unequal weighting effect defined as $1+[(n-1)/n]*CV^2$, where CV=coefficient of variation of weights.

[3] Number of proposed covariates on top line, and number finalized after modeling.

[4] Nominal bounds are used in defining maximum / minimum values for the GEM adjustment factors. The realized bound is the actual adjustment produced by the modeling. The first set of bounds listed is for high extreme values, the second for nonextreme, and the third for low extreme values.

Table H2b 2002 Distribution of Weight Adjustment Factors and Weight Products (Model Group 2: Midwest and West)

	SDU wt	pair selection		sel.pr.ps [1]		res.pr.nr [1]		res.pr.ps [1]		res.pr.ev [1]	
	1-9	pairwt10	1-10	pairwt11	1-11	pairwt12	1-12	pairwt13	1-13	pairwt14	1-14
Minimum	14	1.02	35	0.06	22	0.29	22	0.23	17	0.45	13
1%	97	1.02	141	0.31	104	0.87	107	0.53	77	0.81	68
5%	136	1.10	285	0.39	242	1.00	244	0.63	209	0.88	196
10%	181	1.18	507	0.46	345	1.00	354	0.73	320	0.91	307
25%	445	1.39	869	0.63	774	1.02	798	0.86	771	0.96	759
Median	635	4.00	2,389	0.93	2,200	1.07	2,273	0.98	2,250	0.99	2,251
75%	1,044	10.16	7,006	1.33	6,269	1.21	6,749	1.11	6,781	1.02	6,884
90%	1,634	16.19	15,755	1.71	16,606	1.63	18,743	1.26	18,736	1.06	19,054
95%	1,973	28.65	23,445	1.97	28,378	2.15	35,478	1.36	34,997	1.08	35,722
99%	2,716	49.04	56,660	2.47	73,457	3.70	107,945	1.54	108,908	1.16	109,290
Maximum	6,085	2,720.40	1,689,860	3.63	1,107,367	4.99	1,207,630	1.72	1,214,670	1.59	1,042,480
n	12,432	-	12,432	-	12,432	-	10,033	-	10,033	-	10,033
mean	800	8.83	6,936	1.02	7,260	1.23	8,996	0.99	8,996	0.99	8,996
Max/mean	7.61	-	244	-	153	-	134	-	135	-	116

[1] For a key to modeling abbreviations, see Chapter 7, Exhibit 7.2.

Model Group 2 Overview

Selected Pair–Level Poststratification

All 75 proposed main effects were included in the model.

No two-factor effects involving State/region were kept in the model due to problems with convergence. Out of 125 two-factor effects, 70 were retained in the model.

All three-factor effects were removed from the model due to convergence problems.

Respondent Pair–Level Nonresponse

All proposed factors were retained in the final model.

Respondent Pair–Level Poststratification

All 85 proposed main effects were included in the model.

No two-factor effects involving State/region were kept in the model due to problems with convergence. Out of 125 two-factor effects, 70 were retained in the model.

All three-factor effects were removed from the model due to convergence problems.

Respondent Pair–Level Extreme Value Adjustment

This step used the exactly same variables as in the respondent pair–level poststratification step.

Exhibit H2.1 Covariates for 2002 NSDUH Pair Weights (sel.pr.ps) Model Group 2: Midwest and West

Variables	Level	Proposed	Final	Comments
One-Factor Effects		**75**	**75**	
Intercept	1	1	1	All levels present.
State	26	25	24	All levels present.
Quarter	4	3	3	All levels present.
Population Density	4	3	3	All levels present.
Group Quarter	3	2	2	All levels present.
Household size	3	2	2	All levels present.
Pair Age	15	14	14	All levels present.
Pair Sex	3	2	2	All levels present.
Pair Race	10	9	9	All levels present.
Race of Householder	6	5	5	All levels present.
Rent/housing	5	4	4	All levels present.
Segment % black	3	2	2	All levels present.
Segment % Hispanic	3	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
Two-Factor Effects		**125**	**70**	
Pair Race (5 levels) x Pair Age (6 levels)	5 x 6	20	20	All levels present.
Pair Race (5 levels) x Pair Sex	5 x 3	8	8	All levels present.
Pair Sex x Pair Age (6 levels)	3 x 6	10	10	All levels present.
State/Region x Pair Race (5 levels)	6 x 5	20	0	Drop all; conv.
State/Region x Pair Age (6 levels)	6 x 6	25	0	Drop all; conv.
State/Region x Pair Sex	6 x 3	10	0	Drop all; conv.
Rent/housing x % black	5 x 3	8	8	All levels present.
Rent/housing x % Hispanic	5 x 3	8	8	All levels present.
Rent/housing x % Owner-Occupied	5 x 3	8	8	All levels present.
% Owner-Occupied x % black	3 x 3	4	4	All levels present.
% Owner-Occupied x % Hispanic	3 x 3	4	4	All levels present.
Three-Factor Effects		**12**	**0**	
Pair Race (4 levels) x Pair Sex x Pair Age (3 levels)	4 x 3 x 3	12	0	Drop all; conv.
Total		**212**	**145**	

Exhibit H2.2 Covariates for 2002 NSDUH Pair Weights (res.pr.nr) Model Group 2: Northeast and South

Variables	Level	Proposed	Final	Comments
One-Factor Effects		**75**	**75**	
Intercept	1	1	1	All levels present.
State	26	25	24	All levels present.
Quarter	4	3	3	All levels present.
Population Density	4	3	3	All levels present.
Group Quarter	3	2	2	All levels present.
Household size	3	2	2	All levels present.
Pair Age	15	14	14	All levels present.
Pair Sex	3	2	2	All levels present.
Pair Race	10	9	9	All levels present.
Race of Householder	6	5	5	All levels present.
Rent/housing	5	4	4	All levels present.
Segment % black	3	2	2	All levels present.
Segment % Hispanic	3	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
Two-Factor Effects		**125**	**125**	
Pair Race (5 levels) x Pair Age (6 levels)	5 x 6	20	20	All levels present.
Pair Race (5 levels) x Pair Sex	5 x 3	8	8	All levels present.
Pair Sex x Pair Age (6 levels)	3 x 6	10	10	All levels present.
State/Region x Pair Race (5 levels)	6 x 5	20	20	All levels present.
State/Region x Pair Age (6 levels)	6 x 6	25	25	All levels present.
State/Region x Pair Sex	6 x 3	10	10	All levels present.
Rent/housing x % black	5 x 3	8	8	All levels present.
Rent/housing x % Hispanic	5 x 3	8	8	All levels present.
Rent/housing x % Owner-Occupied	5 x 3	8	8	All levels present.
% Owner-Occupied x % black	3 x 3	4	4	All levels present.
% Owner-Occupied x % Hispanic	3 x 3	4	4	All levels present.
Three-Factor Effects		**12**	**12**	
Pair Race (4 levels) x Pair Sex x Pair Age (3 levels)	4 x 3 x 3	12	12	All levels present.
Total		**212**	**212**	

Exhibit H2.3 Covariates for 2002 NSDUH Pair Weights (res.pr.ps) Model Group 2: Midwest and West

Variables	Level	Proposed	Final	Comments
One-Factor Effects		**85**	**85**	
Intercept	1	1	1	All levels present.
State	26	24	24	All levels present.
Quarter	4	3	3	All levels present.
Population Density	4	3	3	All levels present.
Group Quarter	3	2	2	All levels present.
Household size	3	2	2	All levels present.
Pair Age	15	14	14	All levels present.
Pair Sex	3	2	2	All levels present.
Pair Race	10	9	9	All levels present.
Race of Householder	6	5	5	All levels present.
Rent/housing	5	4	4	All levels present.
Segment % black	3	2	2	All levels present.
Segment % Hispanic	3	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
Pair relationship	10	10	10	All levels present.
Two-Factor Effects		**125**	**70**	
Pair Race (5 levels) x Pair Age (6 levels)	5 x 6	20	20	All levels present.
Pair Race (5 levels) x Pair Sex	5 x 3	8	8	All levels present.
Pair Sex x Pair Age (6 levels)	3 x 6	10	10	All levels present.
State/Region x Pair Race (5 levels)	6 x 5	20	0	Drop all; conv.
State/Region x Pair Age (6 levels)	6 x 6	25	0	Drop all; conv.
State/Region x Pair Sex	6 x 3	10	0	Drop all; conv.
Rent/housing x % black	5 x 3	8	8	All levels present.
Rent/housing x % Hispanic	5 x 3	8	8	All levels present.
Rent/housing x % Owner-Occupied	5 x 3	8	8	All levels present.
% Owner-Occupied x % black	3 x 3	4	4	All levels present.
% Owner-Occupied x % Hispanic	3 x 3	4	4	All levels present.
Three-Factor Effects		**12**	**0**	
Pair Race (4 levels) x Pair Sex x Pair Age (3 levels)	4 x 3 x 3	12	0	Drop all; conv.
Total		**223**	**155**	

Exhibit H2.4 Covariates for 2002 NSDUH Pair Weights (res.pr.ev) Model Group 2: Midwest and West

This step used the same variables as the respondent pair–level poststratification step in Exhibit H2.3.

Appendix I

Evaluation of Calibration Weights: Pair–Level Response Rates

Table I.1 2002 NSDUH Person Pair-Level Response Rates

Domain	Selected Pair Size	Respondent Pair Size	Interview Response Rate [1]
Total	24,895	20,038	71.50%
Pair Age Group			
12-17, 12-17	4,667	4,196	89.75%
12-17, 18-25	3,245	2,748	84.67%
12-17, 26-34	826	689	81.58%
12-17, 35-49	3,795	3,122	80.82%
12-17, 50+	482	370	70.71%
18-25, 18-25	5,520	4,492	80.16%
18-25, 26-34	975	745	74.20%
18-25, 35-49	1,449	1,057	71.70%
18-25, 50+	604	415	68.78%
26-34, 26-34	774	556	69.24%
26-34, 35-49	450	335	77.91%
26-34, 50+	196	119	52.14%
35-49, 35-49	807	539	65.70%
35-49, 50+	350	212	65.44%
50+, 50+	755	443	58.47%
Pair Race			
Hispanic	3,078	2,428	72.42%
black	2,636	2,198	78.92%
white	16,100	12,990	72.10%
other	1,339	1,002	48.75%
white & black	193	153	60.14%
white & Hispanic	726	592	81.11%
white & other	558	462	69.17%
black & Hispanic	80	61	48.17%
black & other	87	73	62.89%
Hispanic & other	98	79	78.34%
Pair Gender			
Male, Male	5,512	4,366	70.57%
Female, Female	5,263	4,432	74.32%
Male, Female	14,120	11,240	70.99%
Household Size			
Two	7,052	5,431	66.04%
Three	6,543	5,176	70.05%
Four or more	11,300	9,431	75.31%

Table I.1 2002 NSDUH Person Pair-Level Response Rates (continued)

Domain	Selected Pairs	Respondent Pairs	Interview Response Rate [1]
Census Region			
Northeast	5,054	3,982	66.60%
South	7,409	6,023	75.42%
Midwest	7,006	5,691	73.10%
West	5,426	4,342	68.64%
Quarter			
Quarter 1	6,012	4,902	71.66%
Quarter 2	5,972	4,813	72.49%
Quarter 3	6,885	5,530	70.68%
Quarter 4	6,026	4,793	71.20%
% Hispanic in Segment			
50-100%	1,493	1,187	76.58%
10-50%	4,203	3,277	64.38%
<10%	19,199	15,574	73.02%
% Black in Segment			
50-100%	1,862	1,554	76.04%
10-50%	3,536	2,821	74.08%
<10%	19,497	15,663	70.48%
% Owner-Occupied DUs in Segment			
50-100%	18,865	15,169	71.92%
10-50%	4,628	3,742	69.93%
<10%	1,402	1,127	68.09%
Combined Median Rent/Housing Value			
1st Quintile	4,594	3,772	75.31%
2nd Quintile	5,030	4,114	75.49%
3rd Quintile	5,313	4,278	71.41%
4th Quintile	4,904	3,892	68.90%
5th Quintile	5,054	3,982	69.02%
Population Density			
Large MSA	8,913	6,978	69.43%
Medium-Small MSA	9,273	7,506	71.78%
Non-MSA, Urban	2,940	2,450	72.30%
Non-MSA, Rural	3,769	3,104	78.10%
Group Quarters			
Group	535	488	92.00%
Nongroup	24,360	19,550	71.40%

[1] The weight used for calculating the response rate includes SDU-level and pair-level design weights, SDU nonresponse and poststratification adjustments, and selected pair poststratification adjustment. This weight is the product of YR02WT1*..*YR02WT9*PR02WT10*PR02WT11.

Appendix J

Evaluation of Calibration Weights: Pair–Level Proportions of Extreme Values and Outwinsors

Table J.1 2002 NSDUH Selected Pair–Level Proportions of Extreme Values and Outwinsors

Domain	*n*	Screener DU-Level Weights (SDUWT: YR02WT1*...*YR02WT9)			Before sel.pr.ps[1] (SDUWT*PR02WT10)			After sel.pr.ps[1] (SDUWT*PR02WT11)		
		Unweighted	Weighted[2]	Outwinsor[3]	Unweighted	Weighted[2]	Outwinsor[3]	Unweighted	Weighted[2]	Outwinsor[3]
Total	24,895	1.60%	3.24%	0.64%	6.05%	24.73%	14.08%	3.01%	16.11%	7.63%
Pair Age Group										
12-17, 12-17	4,667	0.84%	1.82%	0.45%	5.66%	21.21%	7.59%	1.76%	7.93%	1.64%
12-17, 18-25	3,245	1.54%	3.17%	0.80%	10.69%	35.96%	15.39%	3.11%	8.65%	1.19%
12-17, 26-34	826	1.82%	2.38%	0.31%	1.45%	8.98%	3.65%	0.61%	0.47%	0.06%
12-17, 35-49	3,795	1.24%	2.05%	0.44%	2.35%	9.47%	2.96%	0.87%	3.22%	0.66%
12-17, 50+	482	0.83%	2.05%	0.27%	1.87%	5.91%	0.49%	3.11%	12.19%	2.41%
18-25, 18-25	5,520	1.87%	3.77%	0.77%	8.57%	32.58%	15.42%	4.89%	15.72%	2.89%
18-25, 26-34	975	3.49%	7.51%	1.59%	2.15%	9.32%	2.83%	3.28%	11.18%	2.19%
18-25, 35-49	1,449	1.93%	3.46%	0.62%	6.00%	21.12%	6.40%	2.55%	8.29%	1.28%
18-25, 50+	604	1.16%	3.01%	0.38%	2.81%	14.91%	5.13%	1.66%	4.89%	0.34%
26-34, 26-34	774	2.45%	5.16%	1.33%	4.39%	33.09%	20.88%	4.26%	29.00%	12.07%
26-34, 35-49	450	3.11%	5.29%	0.55%	7.56%	32.72%	17.23%	7.56%	36.94%	18.39%
26-34, 50+	196	2.55%	6.27%	0.95%	6.12%	22.87%	7.25%	1.53%	6.78%	2.69%
35-49, 35-49	807	1.86%	3.55%	0.08%	6.20%	52.86%	40.34%	7.06%	51.85%	35.36%
35-49, 50+	350	2.00%	6.75%	0.82%	2.86%	26.81%	20.05%	2.29%	16.15%	9.96%
50+, 50+	755	1.59%	2.01%	0.15%	6.23%	31.79%	23.89%	3.97%	15.29%	6.97%
Pair Race										
Hispanic	3,078	2.63%	5.13%	1.25%	6.63%	25.05%	10.46%	4.13%	17.90%	6.20%
black	2,636	1.97%	3.95%	0.81%	7.63%	22.22%	8.20%	5.88%	18.61%	5.79%
white	16,100	0.58%	1.21%	0.17%	5.43%	24.98%	16.32%	2.10%	15.28%	8.86%
other	1,339	7.47%	14.22%	3.02%	9.71%	28.85%	9.02%	4.33%	17.83%	4.10%
white & black	193	2.59%	3.63%	0.52%	5.70%	13.18%	4.88%	3.11%	5.72%	0.91%
white & Hispanic	726	1.38%	2.19%	0.49%	3.58%	28.03%	22.16%	3.58%	17.94%	10.06%
white & other	558	4.48%	6.10%	0.53%	6.99%	19.18%	4.16%	2.69%	4.95%	1.26%
black & Hispanic	80	16.25%	36.76%	8.75%	8.75%	15.62%	5.04%	20.00%	47.14%	8.26%
black & other	87	9.20%	15.80%	1.23%	5.75%	10.39%	3.74%	1.15%	8.87%	0.89%
Hispanic & other	98	11.22%	25.44%	6.63%	8.16%	27.70%	4.80%	8.16%	13.21%	2.31%
Pair Gender										
Male, Male	5,512	2.36%	5.32%	1.20%	8.96%	26.48%	11.42%	4.34%	12.25%	2.99%
Female, Female	5,263	1.31%	2.23%	0.44%	6.19%	18.50%	6.17%	2.93%	11.47%	3.51%
Male, Female	14,120	1.42%	2.75%	0.48%	4.86%	25.90%	17.04%	2.53%	18.45%	10.04%
Household Size										
Two	7,052	1.23%	2.83%	0.55%	0.51%	1.51%	0.31%	0.51%	2.40%	0.38%
Three	6,543	1.30%	2.23%	0.38%	2.67%	32.53%	25.36%	1.71%	21.64%	13.73%
Four or more	11,300	2.01%	4.02%	0.83%	11.46%	33.44%	15.05%	5.33%	20.71%	8.33%

Table J.1 2002 NSDUH Selected Pair–Level Proportions of Extreme Values and Outwinsors (continued)

Domain	*n*	Screener DU-Level Weights (SDUWT: YR02WT1*...*YR02WT9)			Before sel.pr.ps[1] (SDUWT*PR02WT10)			After sel.pr.ps[1] (SDUWT*PR02WT11)		
		Unweighted	Weighted[2]	Outwinsor[3]	Unweighted	Weighted[2]	Outwinsor[3]	Unweighted	Weighted[2]	Outwinsor[3]
Census Region										
Northeast	5,054	1.86%	4.42%	1.01%	7.22%	18.12%	6.76%	3.60%	14.36%	3.03%
South	7,409	1.28%	2.74%	0.53%	5.28%	26.46%	17.23%	3.36%	19.45%	9.57%
Midwest	7,006	2.00%	3.54%	0.56%	6.14%	28.55%	18.17%	2.54%	18.34%	11.21%
West	5,426	1.29%	2.78%	0.59%	5.90%	23.27%	10.69%	2.60%	10.97%	5.50%
Quarter										
Quarter 1	6,012	1.85%	3.54%	0.75%	7.02%	22.59%	10.47%	4.09%	15.79%	5.18%
Quarter 2	5,972	1.44%	3.38%	0.74%	5.56%	29.72%	20.37%	3.58%	17.55%	8.88%
Quarter 3	6,885	1.13%	2.26%	0.40%	4.92%	22.62%	12.60%	1.96%	15.11%	7.58%
Quarter 4	6,026	2.06%	3.76%	0.68%	6.85%	24.11%	12.95%	2.57%	16.03%	8.88%
% Hispanic in Segment										
50-100%	1,493	2.08%	4.23%	0.96%	6.90%	31.07%	16.90%	4.42%	21.77%	10.67%
10-50%	4,203	2.19%	4.22%	0.90%	6.19%	21.19%	8.10%	4.57%	13.50%	3.72%
<10%	19,199	1.44%	2.84%	0.53%	5.95%	24.88%	15.31%	2.56%	16.20%	8.43%
% Black in Segment										
50-100%	1,862	2.26%	4.95%	0.97%	7.57%	22.15%	7.67%	5.16%	15.60%	5.12%
10-50%	3,536	2.46%	5.53%	1.36%	6.79%	24.01%	11.19%	4.78%	17.38%	6.38%
<10%	19,497	1.38%	2.54%	0.44%	5.77%	25.12%	15.25%	2.49%	15.90%	8.15%
% Owner-Occupied DUs in Segment										
50-100%	18,865	1.36%	2.47%	0.42%	5.76%	24.99%	15.06%	2.85%	16.23%	8.28%
10-50%	4,628	2.33%	5.34%	1.29%	6.94%	23.98%	10.23%	4.43%	16.38%	5.27%
<10%	1,402	2.43%	5.59%	1.19%	6.99%	22.81%	10.45%	0.57%	8.18%	1.30%
Combined Median Rent/Housing Value										
1st Quintile	4,594	1.52%	2.75%	0.47%	6.57%	25.43%	15.86%	2.92%	17.74%	9.08%
2nd Quintile	5,030	1.77%	4.08%	0.88%	6.54%	21.02%	10.49%	2.70%	14.47%	7.41%
3rd Quintile	5,313	1.51%	3.12%	0.65%	6.04%	32.72%	23.25%	2.82%	20.19%	11.84%
4th Quintile	4,904	1.90%	3.33%	0.64%	5.69%	18.54%	6.39%	3.34%	10.12%	2.38%
5th Quintile	5,054	1.33%	2.91%	0.55%	5.44%	24.37%	13.10%	3.28%	18.42%	8.17%
Population Density										
Large MSA	8,913	2.11%	3.94%	0.75%	6.18%	25.81%	14.33%	3.44%	15.40%	6.00%
Medium-Small MSA	9,273	1.48%	3.07%	0.68%	6.15%	19.06%	8.88%	2.90%	14.08%	6.58%
Non-MSA, Urban	2,940	1.05%	2.54%	0.65%	5.61%	19.61%	9.19%	3.64%	9.70%	2.89%
Non-MSA, Rural	3,769	1.14%	1.66%	0.14%	5.84%	35.43%	26.21%	1.78%	28.23%	19.62%
Group Quarters										
Group	535	2.43%	3.18%	0.48%	9.35%	34.30%	15.31%	7.85%	27.74%	6.60%
Nongroup	24,360	1.58%	3.24%	0.64%	5.98%	24.69%	14.07%	2.91%	16.05%	7.63%

[1] This step used demographic variables from screener data for all selected person pairs; Sel = selected, PR = pair, PS = poststratification adjustment.

[2] Weighted extreme value proportion: $100 * \sum_k w_{ek} / \sum_k w_k$, where w_{ek} denotes the weight for extreme values and w_k denotes the weight for both extreme values and non-extreme values.

[3] Outwinsor weight proportion: $100 * \sum_k (w_{ek} - b_k) / \sum_k w_k$, where b_k denotes the winsorized weight.

Table J.2 2002 NSDUH Respondent Pair–Level Proportions of Extreme Values and Outwinsors

Domain	*n*	Before res.pr.nr[1] (SDIWT*PR02WT10*PR02WT11)			After res.pr.nr[1] (SDUWT*PR02WT10*...*PR02WT12)		
		Unweighted	Weighted[2]	Outwinsor[3]	Unweighted	Weighted[2]	Outwinsor[3]
Total	20,038	3.12%	17.80%	8.76%	2.54%	19.16%	8.52%
Pair Age Group							
12-17, 12-17	4,196	1.72%	7.73%	1.59%	0.64%	3.89%	0.93%
12-17, 18-25	2,748	3.24%	8.64%	1.23%	1.49%	5.84%	1.32%
12-17, 26-34	689	0.87%	3.28%	0.17%	0.87%	4.42%	0.31%
12-17, 35-49	3,122	0.96%	4.07%	0.86%	0.86%	3.39%	0.77%
12-17, 50+	370	3.24%	10.42%	1.46%	2.97%	9.51%	2.10%
18-25, 18-25	4,492	5.32%	17.16%	3.30%	3.56%	16.12%	4.28%
18-25, 26-34	745	3.22%	9.99%	2.66%	4.97%	15.17%	6.03%
18-25, 35-49	1,057	2.93%	10.82%	1.47%	5.01%	16.24%	3.08%
18-25, 50+	415	1.93%	7.58%	0.84%	2.41%	4.30%	0.38%
26-34, 26-34	556	4.14%	31.59%	15.32%	5.04%	40.32%	16.28%
26-34, 35-49	335	8.66%	43.54%	23.50%	7.16%	44.95%	17.45%
26-34, 50+	119	0.84%	6.61%	3.80%	5.04%	14.85%	5.05%
35-49, 35-49	539	7.05%	54.73%	38.00%	9.83%	56.50%	33.88%
35-49, 50+	212	3.77%	25.42%	15.75%	3.77%	20.25%	11.16%
50+, 50+	443	3.61%	18.52%	10.65%	4.06%	18.37%	9.29%
Pair Race							
Hispanic	2,428	3.71%	17.58%	6.76%	4.86%	21.31%	7.14%
black	2,198	6.14%	22.92%	7.27%	3.82%	18.47%	5.07%
white	12,990	2.33%	17.19%	10.02%	1.56%	17.40%	8.96%
other	1,002	3.59%	15.02%	3.99%	5.89%	37.87%	15.98%
white & black	153	2.61%	7.80%	0.85%	4.58%	29.97%	10.25%
white & Hispanic	592	3.72%	20.37%	11.91%	1.52%	16.25%	9.56%
white & other	462	3.46%	7.39%	1.94%	1.08%	4.78%	1.02%
black & Hispanic	61	21.31%	56.11%	11.59%	34.43%	43.99%	12.95%
black & other	73	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Hispanic & other	79	8.86%	16.14%	2.65%	3.80%	2.81%	0.64%
Pair Gender							
Male, Male	4,366	4.56%	14.74%	3.97%	3.80%	16.50%	4.63%
Female, Female	4,432	3.02%	11.70%	4.12%	1.90%	15.23%	5.43%
Male, Female	11,240	2.61%	20.38%	11.41%	2.30%	20.96%	10.45%
Household Size							
Two	5,431	0.59%	2.78%	0.51%	0.94%	7.59%	2.26%
Three	5,176	1.68%	24.49%	17.11%	2.05%	27.40%	16.67%
Four or more	9,431	5.38%	21.72%	8.56%	3.73%	21.09%	7.58%

Table J.2 2002 NSDUH Respondent Pair–Level Proportions of Extreme Values and Outwinsors (continued)

Domain	*n*	Before res.pr.nr[1] (SDUWT*PR02WT10*PR02WT11)			After res.pr.nr[1] (SDUWT*PR02WT10*...*PR02WT12)		
		Unweighted	Weighted[2]	Outwinsor[3]	Unweighted	Weighted[2]	Outwinsor[3]
Census Region							
Northeast	3,982	3.64%	16.25%	3.76%	2.71%	17.49%	4.69%
South	6,023	3.64%	22.93%	12.27%	2.69%	20.92%	10.90%
Midwest	5,691	2.71%	19.02%	11.33%	2.30%	21.21%	10.56%
West	4,342	2.49%	10.10%	4.95%	2.49%	16.26%	6.56%
Quarter							
Quarter 1	4,902	4.26%	14.51%	4.75%	2.71%	15.31%	5.22%
Quarter 2	4,813	3.55%	18.78%	9.27%	2.62%	20.93%	9.16%
Quarter 3	5,530	2.12%	18.19%	9.24%	2.30%	19.83%	8.35%
Quarter 4	4,793	2.69%	19.71%	11.78%	2.57%	20.55%	11.37%
% Hispanic in Segment							
50-100%	1,187	4.72%	26.04%	12.91%	3.79%	23.93%	10.13%
10-50%	3,277	4.12%	12.68%	3.15%	5.86%	20.35%	6.72%
<10%	15,574	2.79%	18.09%	9.71%	1.75%	18.20%	8.87%
% Black in Segment							
50-100%	1,554	4.95%	17.64%	6.37%	3.47%	17.05%	4.33%
10-50%	2,821	4.93%	21.26%	8.10%	4.25%	20.02%	7.06%
<10%	15,663	2.62%	17.05%	9.18%	2.14%	19.20%	9.27%
% Owner-Occupied DUs in Segment							
50-100%	15,169	2.91%	18.34%	9.75%	2.25%	19.10%	9.28%
10-50%	3,742	4.76%	16.06%	4.65%	4.14%	20.19%	5.71%
<10%	1,127	0.53%	9.59%	2.94%	1.06%	11.34%	2.14%
Combined Median Rent/Housing Value							
1st Quintile	3,772	3.31%	20.76%	10.80%	1.94%	19.27%	8.84%
2nd Quintile	4,114	2.72%	17.65%	9.76%	1.75%	18.82%	9.17%
3rd Quintile	4,278	2.83%	20.91%	12.52%	3.11%	23.31%	12.03%
4th Quintile	3,892	3.57%	10.35%	2.56%	2.54%	13.31%	3.51%
5th Quintile	3,982	3.24%	20.32%	9.12%	3.31%	21.24%	9.56%
Population Density							
Large MSA	6,978	3.55%	17.98%	8.03%	3.65%	20.70%	8.32%
Medium-Small MSA	7,506	3.09%	14.39%	6.19%	2.17%	15.48%	6.43%
Non-MSA, Urban	2,450	3.71%	8.43%	1.52%	2.61%	10.73%	2.94%
Non-MSA, Rural	3,104	1.77%	30.88%	21.66%	0.87%	28.24%	18.33%
Group Quarters							
Group	488	9.84%	33.79%	7.54%	1.43%	9.53%	3.33%
Nongroup	19,550	2.96%	17.69%	8.77%	2.57%	19.20%	8.55%

[1] This step used demographic variables from screener data for all responding person pairs; Res = respondent, PR = pair, NR = nonresponse adjustment.
[2] Weighted extreme value proportion: $100 * \sum_k w_{ek}/\sum_k w_k$, where w_{ek} denotes the weight for extreme values and w_k denotes the weight for both extreme values and non-extreme values.
[3] Outwinsor weight proportion: $100 * \sum_k (w_{ek}-b_k)/\sum_k w_k$, where b_k denotes the winsorized weight.

Table J.3 2002 NSDUH Respondent Pair–Level Proportions of Extreme Values and Outwinsors

Domain	*n*	Before res.pr.ps[1] (SDUWT*PR02WT10*...*PR02WT12)			After res.pr.ps[1] (SDUWT*PR02WT10*...*PR02WT13)			Final Weight: After res.pr.ev[1] (SDUWT*PR02WT10*...*PR02WT14)		
		Unweighted	Weighted[2]	Outwinsor[3]	Unweighted	Weighted[2]	Outwinsor[3]	Unweighted	Weighted[2]	Outwinsor[3]
Total	20,038	2.40%	15.04%	4.16%	2.06%	11.09%	1.95%	0.76%	3.81%	0.46%
Pair Age Group										
12-17, 12-17	4,192	0.64%	4.02%	1.05%	0.74%	3.31%	0.37%	0.19%	1.47%	0.17%
12-17, 18-25	2,742	1.57%	5.77%	1.54%	1.50%	4.92%	0.45%	0.44%	1.38%	0.07%
12-17, 26-34	694	0.86%	6.81%	0.65%	1.15%	4.37%	0.28%	0.43%	3.19%	0.20%
12-17, 35-49	3,121	1.12%	3.99%	1.06%	0.93%	3.19%	0.37%	0.22%	0.79%	0.06%
12-17, 50+	377	0.80%	2.98%	0.29%	0.27%	0.27%	0.06%	0.00%	0.00%	0.00%
18-25, 18-25	4,419	3.64%	16.57%	4.44%	2.96%	11.11%	1.66%	0.91%	4.04%	0.23%
18-25, 26-34	806	4.96%	16.36%	5.94%	3.72%	12.28%	3.37%	2.36%	7.88%	0.68%
18-25, 35-49	1,042	4.22%	14.47%	2.87%	3.93%	10.92%	1.81%	1.44%	3.58%	0.24%
18-25, 50+	418	0.72%	0.72%	0.14%	0.72%	2.37%	0.74%	0.48%	0.61%	0.15%
26-34, 26-34	559	3.04%	20.67%	7.33%	2.68%	12.20%	2.43%	0.72%	2.80%	0.67%
26-34, 35-49	346	6.65%	16.55%	3.67%	4.05%	14.97%	3.49%	4.05%	9.29%	1.47%
26-34, 50+	123	9.76%	27.66%	4.40%	4.07%	11.38%	1.00%	0.00%	0.00%	0.00%
35-49, 35-49	543	7.00%	35.55%	11.42%	4.97%	23.52%	3.57%	1.84%	4.42%	1.02%
35-49, 50+	210	3.81%	20.16%	3.09%	3.81%	20.23%	4.22%	0.95%	5.68%	1.27%
50+, 50+	446	4.71%	19.94%	6.91%	6.50%	16.39%	3.08%	3.81%	8.50%	0.49%
Pair Race										
Hispanic	2,464	4.38%	15.79%	3.39%	3.13%	8.71%	1.30%	0.85%	2.00%	0.08%
black	2,199	3.82%	11.67%	2.62%	2.96%	7.25%	1.26%	0.68%	2.37%	0.17%
white	12,878	1.38%	12.10%	3.29%	1.26%	9.36%	1.34%	0.33%	2.02%	0.12%
other	922	6.18%	39.75%	15.46%	6.94%	45.10%	11.25%	4.99%	27.06%	5.18%
white & black	168	5.95%	25.35%	6.84%	7.74%	22.60%	6.28%	3.57%	9.75%	1.48%
white & Hispanic	569	1.05%	4.88%	2.38%	2.64%	5.51%	1.30%	1.93%	4.90%	0.49%
white & other	548	2.55%	38.41%	9.57%	0.36%	1.56%	0.10%	0.00%	0.00%	0.00%
black & Hispanic	74	21.62%	44.44%	12.87%	16.22%	37.02%	6.25%	14.86%	35.44%	3.11%
black & other	103	1.94%	5.66%	1.36%	2.91%	3.18%	0.45%	0.00%	0.00%	0.00%
Hispanic & other	113	5.31%	7.75%	1.08%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Pair Gender										
Male, Male	4,366	3.66%	14.43%	3.67%	2.86%	7.88%	1.23%	1.10%	2.90%	0.33%
Female, Female	4,430	1.72%	10.30%	3.32%	1.63%	9.30%	2.25%	0.72%	5.06%	0.44%
Male, Female	11,242	2.18%	16.49%	4.52%	1.92%	12.48%	2.06%	0.65%	3.72%	0.50%
Household Size										
Two	5,431	0.90%	7.55%	2.20%	1.10%	8.79%	1.94%	0.72%	5.25%	0.78%
Three	5,176	2.20%	27.08%	8.29%	2.11%	20.18%	3.14%	1.06%	5.94%	0.64%
Four or more	9,431	3.37%	12.65%	3.00%	2.59%	7.44%	1.30%	0.63%	1.85%	0.18%

Table J.3 2002 NSDUH Respondent Pair–Level Proportions of Extreme Values and Outwinsors (continued)

Domain	*n*	Before res.pr.ps[1] (SDUWT*PR02WT10*...*PR02WT12)			After res.pr.ps[1] (SDUWT*PR02WT10*...*PR02WT13)			Final Weight: After res.pr.ev[1] (SDUWT*PR02WT10*...*PR02WT14)		
		Unweighted	Weighted[2]	Outwinsor[3]	Unweighted	Weighted[2]	Outwinsor[3]	Unweighted	Weighted[2]	Outwinsor[3]
Census Region										
Northeast	3,982	2.76%	16.56%	4.28%	2.56%	15.23%	3.56%	1.48%	8.18%	1.18%
South	6,023	2.42%	16.46%	6.10%	2.44%	10.87%	2.13%	1.08%	4.66%	0.52%
Midwest	5,691	2.07%	13.18%	2.04%	1.41%	9.86%	0.89%	0.14%	0.24%	0.02%
West	4,342	2.46%	13.48%	3.25%	1.93%	9.14%	1.32%	0.48%	2.26%	0.17%
Quarter										
Quarter 1	4,902	2.73%	14.36%	3.47%	2.22%	11.60%	2.37%	1.00%	4.90%	0.75%
Quarter 2	4,813	2.43%	18.06%	5.87%	2.49%	10.31%	1.91%	0.83%	4.85%	0.44%
Quarter 3	5,530	2.06%	11.32%	2.36%	1.54%	9.76%	1.50%	0.45%	2.31%	0.24%
Quarter 4	4,793	2.42%	16.55%	4.99%	2.07%	12.71%	2.02%	0.81%	3.21%	0.41%
% Hispanic in Segment										
50-100%	1,187	3.29%	10.13%	1.99%	2.70%	6.13%	0.87%	0.67%	1.50%	0.08%
10-50%	3,277	5.86%	21.57%	5.95%	4.58%	14.79%	2.78%	1.62%	5.07%	0.55%
<10%	15,574	1.61%	13.67%	3.88%	1.48%	10.59%	1.83%	0.59%	3.71%	0.47%
% Black in Segment										
50-100%	1,554	3.15%	10.49%	2.21%	3.35%	6.93%	1.02%	0.77%	3.02%	0.11%
10-50%	2,821	4.11%	17.12%	4.20%	3.15%	8.34%	1.88%	1.35%	4.73%	0.53%
<10%	15,663	2.02%	15.09%	4.35%	1.74%	12.11%	2.06%	0.66%	3.70%	0.48%
% Owner-Occupied DUs in Segment										
50-100%	15,169	2.08%	14.79%	4.12%	1.73%	10.55%	1.87%	0.57%	3.15%	0.39%
10-50%	3,742	4.09%	16.55%	4.54%	3.77%	13.93%	2.44%	1.66%	6.67%	0.74%
<10%	1,127	1.15%	11.53%	2.14%	0.89%	7.26%	0.55%	0.44%	5.15%	0.60%
Combined Median Rent/Housing Value										
1st Quintile	3,772	1.83%	10.96%	2.69%	1.54%	7.47%	1.48%	0.50%	3.07%	0.30%
2nd Quintile	4,114	1.43%	11.92%	4.63%	2.09%	12.41%	1.80%	0.73%	3.06%	0.29%
3rd Quintile	4,278	2.90%	17.22%	5.70%	2.22%	8.45%	1.36%	0.84%	2.14%	0.24%
4th Quintile	3,892	2.62%	14.49%	3.67%	2.24%	10.70%	2.01%	0.72%	4.16%	0.31%
5th Quintile	3,982	3.19%	18.22%	3.75%	2.18%	15.19%	2.82%	1.00%	5.99%	1.02%
Population Density										
Large MSA	6,978	3.48%	16.20%	4.89%	2.68%	10.95%	2.14%	0.95%	4.21%	0.58%
Medium-Small MSA	7,506	2.12%	13.35%	4.17%	2.10%	11.96%	2.07%	0.89%	4.77%	0.52%
Non-MSA, Urban	2,450	2.37%	7.87%	2.10%	1.96%	5.12%	0.62%	0.45%	1.54%	0.05%
Non-MSA, Rural	3,104	0.68%	19.67%	2.71%	0.64%	13.33%	1.76%	0.29%	1.29%	0.09%
Group Quarters										
Group	488	1.43%	9.97%	3.36%	2.66%	17.05%	3.89%	3.28%	19.77%	1.91%
Nongroup	19,550	2.42%	15.06%	4.16%	2.05%	11.06%	1.94%	0.70%	3.73%	0.45%

Table J.3 2002 NSDUH Respondent Pair–Level Proportions of Extreme Values and Outwinsors (continued)

Domain	*n*	Before res.pr.ps[1] (SDUWT*PR02WT10*...*PR02WT12)			After res.pr.ps[1] (SDUWT*PR02WT10*...*PR02WT13)			Final Weight: After res.pr.ev[1] (SDUWT*PR02WT10*...*PR02WT14)		
		Unweighted	Weighted[2]	Outwinsor[3]	Unweighted	Weighted[2]	Outwinsor[3]	Unweighted	Weighted[2]	Outwinsor[3]
Pair Relationship[4]										
Parent-child (12-14)	2,133	1.08%	3.72%	0.70%	0.98%	3.07%	0.43%	0.23%	0.86%	0.05%
Parent-child (12-17)	3,864	0.98%	3.58%	0.86%	1.01%	2.72%	0.33%	0.26%	0.65%	0.05%
Parent-child (12-20)	4,545	1.43%	5.51%	1.27%	1.72%	5.15%	0.79%	0.51%	1.45%	0.10%
Sibling (12-14) - sibling (15-17)	2,464	0.49%	2.70%	0.63%	0.65%	2.42%	0.20%	0.08%	0.32%	0.01%
Sibling (12-17) - sibling (18-25)	2,425	1.48%	5.92%	1.67%	1.44%	4.93%	0.45%	0.45%	1.56%	0.08%
Spouse-spouse	4,000	2.08%	21.82%	6.58%	2.95%	18.49%	3.35%	1.23%	6.22%	0.87%
Spouse-spouse with children (under 18)	2,020	1.98%	25.92%	8.01%	3.61%	23.03%	4.67%	1.24%	5.86%	1.35%

[1] This step used demographic variables from questionnaire data for all responding person pairs; Res = respondent, PR = pair, PS = poststratification adjustment, EV = extreme value adjustment.

[2] Weighted extreme value proportion: $100 * \sum_k w_{ek}/\sum_k w_k$, where w_{ek} denotes the weight for extreme values and w_k denotes the weight for both extreme values and non-extreme values.

[3] Outwinsor weight proportion: $100 * \sum_k (w_{ek}-b_k)/\sum_k w_k$, where b_k denotes the winsorized weight.

[4] Parent-child (15-17) was not included here since extreme values were not controlled with this domain. Spouse-spouse pair relationships also included partner-partner relationships.

Appendix K

Evaluation of Calibration Weights: Pair-Level Slippage Rates

Table K.1 2002 NSDUH Respondent Pair-Level Slippage Rates

Domain	*n*	Initial Total (I)[1]	Final Total (F)[2]	Control Total from SDU (C)	(I-C)/C %	(F-C)/C %
Total	20,038	194,935,867	194,935,867	194,935,867	0.00	-0.00
Pair Age Group						
12-17, 12-17	4,192	7,792,895	7,745,463	7,745,463	0.61	0.00
12-17, 18-25	2,742	7,403,004	7,481,518	7,481,518	-1.05	0.00
12-17, 26-34	694	5,366,972	5,163,413	5,163,413	3.94	0.00
12-17, 35-49	3,121	32,346,605	32,540,796	32,540,796	-0.60	0.00
12-17, 50+	377	8,204,008	8,019,369	8,019,369	2.30	0.00
18-25, 18-25	4,419	11,062,901	11,117,904	11,117,904	-0.49	0.00
18-25, 26-34	806	6,989,972	6,997,837	6,997,837	-0.11	0.00
18-25, 35-49	1,042	16,510,668	16,837,725	16,837,725	-1.94	0.00
18-25, 50+	418	10,806,065	10,829,874	10,829,874	-0.22	0.00
26-34, 26-34	559	9,930,017	10,185,649	10,185,649	-2.51	-0.00
26-34, 35-49	346	9,118,132	8,459,446	8,459,446	7.79	-0.00
26-34, 50+	123	7,529,345	7,583,682	7,583,682	-0.72	-0.00
35-49, 35-49	543	20,633,741	20,386,940	20,386,940	1.21	-0.00
35-49, 50+	210	12,842,197	13,971,572	13,971,572	-8.08	-0.00
50+, 50+	446	28,399,346	27,614,678	27,614,678	2.84	-0.00
Pair Race						
Hispanic	2,464	28,982,740	28,420,414	28,420,414	1.98	0.00
black	2,199	22,167,942	21,644,011	21,644,011	2.42	-0.00
white	12,878	118,001,749	120,032,021	120,032,021	-1.69	0.00
other	922	11,064,063	11,703,604	11,703,604	-5.46	-0.00
white & black	168	1,293,082	1,462,084	1,462,084	-11.56	0.00
white & Hispanic	569	4,886,107	5,756,876	5,756,876	-15.13	0.00
white & other	548	6,192,585	3,816,560	3,816,560	62.26	0.00
black & Hispanic	74	984,057	914,192	914,192	7.64	-0.00
black & other	103	595,239	604,402	604,402	-1.52	0.00
Hispanic & other	113	768,304	581,703	581,703	32.08	0.00
Pair Gender						
Male, Male	4,366	35,327,475	35,263,303	35,263,303	0.18	-0.00
Female, Female	4,430	33,928,378	33,992,862	33,992,862	-0.19	-0.00
Male, Female	11,242	125,680,014	125,679,702	125,679,702	0.00	-0.00
Pair Domain[3,4,5]						
Parent-child (12-14) *	365	12,158,933	12,691,380	12,691,380	-4.20	0.00
Parent-child (12-17)*	687	23,047,430	24,423,926	24,423,926	-5.64	0.00
Parent-child (15-17)*	324	10,890,051	11,733,587	11,732,546	-7.18	0.01
Parent-child (12-20)*	824	30,049,758	32,406,091	32,406,091	-7.27	0.00
Parent*-child (12-14)	1,862	18,474,117	19,442,193	19,442,193	-4.98	0.00
Parent*-child (12-17)	2,733	29,032,614	31,143,460	31,143,460	-6.78	0.00
Parent*-child (15-17)	1,521	17,243,481	18,404,019	18,080,125	-4.63	1.79
Parent*-child (12-20)	2,894	34,981,434	37,801,678	37,801,678	-7.46	0.00
Sibling (12-14) -sibling (15-17)*	2,243	3,994,532	4,105,623	4,105,623	-2.71	0.00
Sibling (12-17) -sibling (18-25)*	1,957	5,188,096	5,175,903	5,175,903	0.24	0.00
Spouse-spouse/partner-partner	4,000	63,447,244	65,710,948	65,710,948	-3.44	-0.00
Spouse-spouse/partner-partner with children under 18	2,020	25,543,252	29,061,451	29,061,451	-12.11	-0.00

Table K.1 2002 NSDUH Respondent Pair-Level Slippage Rates (continued)

Domain	*n*	Initial Total (I)[1]	Final Total (F)[2]	Control Total from SDU (C)	(I-C)/C %	(F-C)/C %
Household Size						
Two	5,431	51,530,620	51,530,620	51,530,620	-0.00	-0.00
Three	5,176	50,456,838	50,456,838	50,456,838	0.00	-0.00
Four or more	9,431	92,948,409	92,948,409	92,948,409	0.00	-0.00
Census Region						
Northeast	3,982	38,686,504	38,686,504	38,686,504	0.00	0.00
South	6,023	65,994,610	65,994,610	65,994,610	0.00	0.00
Midwest	5,691	42,287,950	42,287,950	42,287,950	-0.00	-0.00
West	4,342	47,966,803	47,966,803	47,966,803	-0.00	-0.00
Quarter						
Quarter 1	4,902	48,487,558	48,487,558	48,487,558	0.00	-0.00
Quarter 2	4,813	48,090,470	48,090,470	48,090,470	0.00	-0.00
Quarter 3	5,530	49,807,019	49,807,019	49,807,019	0.00	-0.00
Quarter 4	4,793	48,550,820	48,550,820	48,550,820	0.00	-0.00
% Hispanic in Segment						
50-100%	1,187	16,927,526	16,927,526	16,927,526	-0.00	-0.00
10-50%	3,277	41,304,753	41,304,753	41,304,753	0.00	-0.00
<10%	15,574	136,703,587	136,703,587	136,703,587	0.00	-0.00
% Black in Segment						
50-100%	1,554	15,725,006	15,725,006	15,725,006	0.00	0.00
10-50%	2,821	30,997,705	30,997,705	30,997,705	0.00	-0.00
<10%	15,663	148,213,156	148,213,156	148,213,156	0.00	-0.00
% Owner-Occupied DUs in Segment						
50-100%	15,169	157,175,431	157,175,430	157,175,430	0.00	-0.00
10-50%	3,742	34,321,840	34,321,840	34,321,840	0.00	-0.00
<10%	1,127	3,438,597	3,438,597	3,438,597	-0.00	-0.00
Combined Median Rent/Housing Value						
1st Quintile	3,772	27,704,709	27,704,709	27,704,709	0.00	-0.00
2nd Quintile	4,114	32,689,473	32,689,473	32,689,473	0.00	-0.00
3rd Quintile	4,278	42,997,335	42,997,335	42,997,335	-0.00	-0.00
4th Quintile	3,892	45,548,030	45,548,030	45,548,030	0.00	-0.00
5th Quintile	3,982	45,996,320	45,996,320	45,996,320	0.00	-0.00
Population Density						
Large MSA	6,978	92,404,772	92,404,772	92,404,772	0.00	-0.00
Medium-Small MSA	7,506	62,227,497	62,227,496	62,227,497	0.00	-0.00
Non-MSA, Urban	2,450	15,988,588	15,988,588	15,988,588	0.00	0.00
Non-MSA, Rural	3,104	24,315,011	24,315,011	24,315,011	0.00	-0.00
Group Quarters						
Group	488	969,960	969,960	969,960	-0.00	0.00
Nongroup	19,550	193,965,908	193,965,907	193,965,907	0.00	-0.00

[1] YR02WT1*...*YR02WT9*PR02WT10*...*PR02WT12 (before person pair poststratification).

[2] YR02WT1*...*YR02WT9*PR02WT10*...*PR02WT13 (after person pair poststratification).

[3] The member of pair that is the focus is designated with an *.

[4] The parent-child (15-17) pair domains were not controlled for within the modeling and thus have higher slippage rates than the other domains listed. However , since these domains are a subset of other controlled domains, the rates are not large.

[5] Slippage rates were not calculated for the sibling-sibling domains with the younger child as the focus since no household counts for this domain were calculated and are required to construct the appropriate controls totals.

Appendix L

Evaluation of Calibration Weights: Pair-Level Weight Summary Statistics

Table L.1 2002 NSDUH Selected Pair–Level Weight Summary Statistics

Domain	*n*	Screener DU-Level Weights (SDUWT: YR02WT1*...*YR02WT9)						Before sel.pr.ps[1] (SDUWT*PR02WT10)						After sel.pr.ps[1] (SDUWT*PR02WT11)					
		Min	Q1[2]	Med	Q3[2]	Max	UWE[3]	Min	Q1[2]	Med	Q3[2]	Max	UWE[3]	Min	Q1[2]	Med	Q3[2]	Max	UWE[3]
Total	24,895	14	490	720	1,143	8,832	1.47	25	983	2,560	7,603	5,413,347	33.99	6	875	2,462	7,101	1,575,006	12.21
Pair Age Group																			
12-17, 12-17	4,667	20	420	646	1,030	5,647	1.53	25	630	1,025	1,945	47,959	2.85	6	440	939	2,019	24,568	2.67
12-17, 18-25	3,245	20	523	739	1,164	7,462	1.46	44	759	1,265	2,426	48,206	3.15	22	684	1,424	2,790	24,698	2.41
12-17, 26-34	826	20	488	693	1,103	2,978	1.44	106	2,671	4,347	7,104	99,645	2.62	170	1,895	3,768	7,235	81,311	2.71
12-17, 35-49	3,795	14	448	686	1,091	7,910	1.47	170	3,456	5,963	10,248	257,307	2.50	104	2,452	4,948	10,167	145,313	2.67
12-17, 50+	482	46	407	727	1,142	3,511	1.45	939	6,459	11,510	18,879	108,498	1.89	441	4,634	10,259	20,998	155,112	2.36
18-25, 18-25	5,520	30	543	758	1,162	8,832	1.44	37	718	1,154	2,091	89,662	3.84	27	481	1,020	2,362	26,294	2.83
18-25, 26-34	975	58	569	809	1,282	5,069	1.43	298	3,030	4,748	7,792	153,587	2.92	122	1,730	3,730	7,675	151,982	3.64
18-25, 35-49	1,449	63	511	756	1,262	5,498	1.44	572	3,959	7,050	12,351	275,476	2.98	178	2,961	6,148	14,001	138,829	2.75
18-25, 50+	604	65	571	860	1,290	5,184	1.42	1,044	8,103	12,842	22,083	236,425	2.52	484	5,896	11,211	22,149	132,451	2.16
26-34, 26-34	774	45	546	754	1,195	7,332	1.45	454	5,402	8,176	13,161	1,248,498	14.68	288	3,394	6,833	11,822	575,039	6.55
26-34, 35-49	450	48	497	739	1,257	3,256	1.41	998	6,064	9,522	16,876	452,426	5.32	733	4,642	9,448	17,731	732,421	7.25
26-34, 50+	196	39	542	823	1,231	3,939	1.45	1,318	13,138	22,731	39,612	418,550	2.42	401	9,001	23,511	52,341	261,553	2.33
35-49, 35-49	807	44	482	728	1,135	3,203	1.43	712	6,909	10,511	16,391	1,513,028	18.33	369	5,262	11,220	19,638	1,572,402	14.03
35-49, 50+	350	60	413	701	1,123	3,694	1.49	2,177	11,160	17,764	29,676	1,689,857	11.21	700	11,525	24,084	47,903	1,107,367	4.14
50+, 50+	755	74	468	711	1,149	2,852	1.40	2,465	15,478	23,099	35,435	5,413,347	33.71	1,445	15,302	27,652	43,135	1,575,006	3.93
Pair Race																			
Hispanic	3,078	14	539	872	1,372	7,910	1.44	35	1,207	3,019	8,626	418,550	5.81	11	1,025	2,919	8,373	732,421	7.34
black	2,636	49	605	848	1,208	5,949	1.36	57	1,150	2,951	7,745	239,249	4.46	12	968	2,747	7,489	396,867	6.65
white	16,100	25	491	693	1,070	5,498	1.42	67	958	2,459	7,414	5,413,347	48.52	24	880	2,367	6,798	1,575,006	15.71
other	1,339	20	232	568	1,297	8,832	1.96	25	777	2,369	6,813	236,425	5.90	7	625	2,410	7,340	231,614	6.29
white & black	193	81	618	866	1,311	3,744	1.38	94	1,180	2,394	6,835	99,429	4.27	59	1,037	2,299	5,841	152,664	5.55
white & Hispanic	726	20	470	714	1,175	4,739	1.48	85	905	2,581	8,326	1,248,498	35.32	31	704	2,512	8,452	575,039	10.26
white & other	558	33	263	568	938	4,116	1.74	44	908	2,204	6,077	158,045	5.67	31	741	1,883	5,391	213,374	8.39
black & Hispanic	80	55	502	818	1,423	6,085	1.87	99	922	2,584	11,819	88,203	3.63	21	1,521	4,468	13,728	128,443	4.17
black & other	87	20	448	704	1,102	3,546	1.52	31	821	1,797	8,670	100,492	4.54	6	510	1,393	4,755	76,998	5.34
Hispanic & other	98	23	162	519	1,124	4,170	2.05	75	630	1,560	4,173	54,011	4.60	85	575	1,598	4,877	61,728	4.78
Pair Gender																			
Male, Male	5,512	22	509	747	1,184	7,910	1.51	54	935	2,267	6,522	586,819	6.70	11	832	2,287	6,280	351,344	5.86
Female, Female	5,263	20	485	700	1,109	5,949	1.46	25	958	2,397	6,582	418,550	5.64	19	821	2,264	6,026	396,867	6.42
Male, Female	14,120	14	485	718	1,138	8,832	1.45	29	1,018	2,812	8,329	5,413,347	45.16	6	922	2,628	8,021	1,575,006	14.22
Household Size																			
Two	7,052	30	496	714	1,106	8,832	1.42	37	948	2,728	9,230	83,678	2.93	31	625	1,887	7,718	157,780	4.24
Three	6,543	22	492	720	1,119	5,992	1.42	25	1,131	3,412	7,063	5,413,347	94.65	11	1,066	2,898	6,577	1,575,006	24.57
Four or more	11,300	14	483	725	1,182	7,910	1.51	29	952	2,181	7,053	1,026,298	9.43	6	986	2,525	7,197	1,110,643	9.79

Table L.1 2002 NSDUH Selected Pair–Level Weight Summary Statistics (continued)

Domain	*n*	Screener DU-Level Weights (SDUWT: YR02WT1*...*YR02WT9)						Before sel.pr.ps[1] (SDUWT*PR02WT10)						After sel.pr.ps[1] (SDUWT*PR02WT11)					
		Min	Q1[2]	Med	Q3[2]	Max	UWE[3]	Min	Q1[2]	Med	Q3[2]	Max	UWE[3]	Min	Q1[2]	Med	Q3[2]	Max	UWE[3]
Census Region																			
Northeast	5,054	20	384	683	968	7,910	1.47	25	891	2,175	7,276	257,307	4.56	7	850	2,347	7,386	396,867	5.41
South	7,409	20	666	953	1,315	8,832	1.33	31	1,331	3,353	9,120	5,413,347	57.37	6	1,127	3,065	8,244	1,575,006	15.50
Midwest	7,006	20	515	623	801	5,069	1.30	54	833	2,218	6,430	1,689,857	29.02	31	776	2,012	5,194	1,107,367	17.35
West	5,426	14	275	688	1,472	6,085	1.68	35	966	2,551	8,190	1,248,498	10.60	22	770	2,463	7,935	732,421	8.24
Quarter																			
Quarter1	6,012	20	499	733	1,201	7,332	1.46	29	977	2,587	8,177	1,190,368	10.47	7	857	2,577	7,845	783,080	7.51
Quarter2	5,972	20	497	712	1,126	7,910	1.45	31	943	2,388	7,164	5,413,347	91.93	6	955	2,516	7,220	1,575,006	15.25
Quarter3	6,885	14	469	673	1,035	5,684	1.42	35	954	2,467	7,124	1,248,498	13.82	21	826	2,271	6,453	732,421	10.31
Quarter4	6,026	20	509	776	1,262	8,832	1.49	25	1,071	2,812	7,964	1,689,857	18.89	12	884	2,518	7,063	1,572,402	15.54
% Hispanic in Segment																			
50-100%	1,493	43	699	1,115	1,492	7,910	1.30	54	1,661	4,105	11,118	1,248,498	13.19	11	1,199	3,562	10,159	732,421	8.88
10-50%	4,203	22	584	945	1,459	7,462	1.39	25	1,322	3,160	9,265	601,924	5.37	17	1,159	3,362	9,235	397,074	5.30
<10%	19,199	14	459	678	1,025	8,832	1.47	29	917	2,320	7,117	5,413,347	46.24	6	810	2,255	6,412	1,575,006	15.19
% Black in Segment																			
50-100%	1,862	20	568	805	1,167	8,832	1.44	31	1,099	2,965	8,132	151,182	4.25	6	955	2,695	7,794	396,867	6.40
10-50%	3,536	20	593	880	1,336	7,462	1.42	25	1,112	2,825	8,018	1,026,298	9.74	11	998	2,811	8,138	1,110,643	9.29
<10%	19,497	14	461	689	1,099	7,910	1.47	29	952	2,477	7,446	5,413,347	41.85	7	853	2,385	6,834	1,575,006	13.55
% Owner-Occupied DUs in Segment																			
50-100%	18,865	14	482	699	1,109	8,832	1.45	29	1,002	2,740	7,953	5,413,347	39.51	6	988	2,712	7,651	1,575,006	12.83
10-50%	4,628	20	519	784	1,226	7,910	1.49	25	956	2,263	6,756	418,550	6.45	12	828	2,255	6,789	559,184	6.97
<10%	1,402	20	567	823	1,290	6,085	1.47	54	903	1,821	5,336	220,322	5.36	11	340	785	2,264	136,298	8.44
Combined Median Rent/Housing Value																			
1st Quintile	4,594	23	372	608	841	4,170	1.49	49	804	1,961	5,707	1,190,368	23.25	22	724	1,911	5,201	1,110,643	16.97
2nd Quintile	5,030	14	320	631	978	8,832	1.65	35	814	2,084	6,521	1,513,028	16.32	32	739	2,107	6,064	1,572,402	17.23
3rd Quintile	5,313	22	541	763	1,129	7,462	1.40	25	1,081	2,793	7,812	5,413,347	86.50	9	927	2,472	6,944	1,575,006	17.40
4th Quintile	4,904	22	606	902	1,325	5,647	1.36	29	1,216	3,212	9,314	236,425	4.26	7	1,123	3,202	9,039	280,277	4.87
5th Quintile	5,054	20	555	787	1,282	7,910	1.41	31	1,148	3,071	8,467	1,689,857	15.17	6	986	2,841	8,282	1,107,367	9.63
Population Density																			
Large MSA	8,913	20	646	906	1,382	7,910	1.32	31	1,383	3,770	10,185	5,413,347	40.96	6	1,332	3,601	10,045	1,575,006	8.03
Medium-Small MSA	9,273	14	412	666	1,020	8,832	1.49	25	908	2,224	6,592	1,513,028	11.42	7	796	2,201	6,071	1,572,402	12.77
Non-MSA, Urban	2,940	31	259	587	930	5,992	1.60	49	752	1,688	4,920	601,924	9.89	14	632	1,765	4,851	397,074	6.63
Non-MSA, Rural	3,769	20	253	589	919	3,917	1.60	56	814	2,106	6,234	1,689,857	38.73	22	581	1,721	4,861	1,110,643	32.13
Group Quarters																			
Group	535	47	460	666	1,032	2,876	1.44	106	687	945	1,571	17,864	3.04	69	494	866	1,748	25,997	3.43
Nongroup	24,360	14	490	722	1,147	8,832	1.47	25	1,003	2,652	7,752	5,413,347	33.58	6	901	2,522	7,263	1,575,006	12.07

[1] This step used demographic variables from screener data for all selected person pairs; Sel = selected, PR = pair, PS = poststratification.
[2] Q1 and Q3 refer to the first and third quartile of the weight distribution.
[3] Unequal weighting effect defined as $1+[(n-1)/n]*CV^2$, where CV=coefficient of variation of weights.

Table L.2 2002 NSDUH Respondent Pair–Level Weight Summary Statistics

Domain	*n*	Before res.pr.nr[1] (SDUWT*PR02WT10*PR02WT11)						After res.pr.nr[1] (SDUWT*PR02WT10*...*PR02WT12)					
		Min	Q1[2]	Med	Q3[2]	Max	UWE[3]	Min	Q1[2]	Med	Q3[2]	Max	UWE[3]
Total	20,038	6	813	2,237	6,238	1,575,006	15.27	6	905	2,601	7,661	2,072,832	14.73
Pair Age Group													
12-17, 12-17	4,196	6	438	937	2,018	24,568	2.67	6	477	1,063	2,315	40,915	2.67
12-17, 18-25	2,748	22	682	1,429	2,789	20,503	2.37	22	774	1,657	3,258	39,256	2.48
12-17, 26-34	689	170	1,848	3,559	7,048	61,504	2.72	170	2,147	4,202	8,458	94,445	3.00
12-17, 35-49	3,122	104	2,419	4,900	9,958	145,313	2.71	104	2,714	5,723	12,109	178,472	2.91
12-17, 50+	370	441	4,270	9,857	20,043	126,197	2.25	441	5,165	13,764	28,231	160,730	2.34
18-25, 18-25	4,492	27	473	996	2,271	26,294	2.86	27	511	1,123	2,849	51,271	3.22
18-25, 26-34	745	122	1,621	3,531	7,433	151,982	4.04	122	1,818	4,034	9,277	176,626	4.22
18-25, 35-49	1,057	232	2,833	6,060	13,662	114,659	2.74	232	3,478	7,479	18,572	246,940	2.92
18-25, 50+	415	512	5,920	11,705	21,694	132,451	2.21	512	7,722	15,120	34,104	177,052	2.32
26-34, 26-34	556	352	3,332	6,603	11,120	575,039	8.07	354	3,761	8,005	14,585	623,207	7.34
26-34, 35-49	335	733	4,571	9,245	16,865	732,421	8.37	733	4,943	10,409	20,628	934,741	8.16
26-34, 50+	119	401	8,748	17,622	44,241	261,553	2.38	519	11,900	33,455	83,530	419,424	2.56
35-49, 35-49	539	433	5,262	10,598	18,928	1,572,402	16.96	489	8,067	16,502	29,417	2,006,610	12.61
35-49, 50+	212	700	10,539	24,275	47,064	1,107,367	5.05	718	11,974	34,988	77,388	1,207,626	3.76
50+, 50+	443	1,972	13,232	26,606	42,092	1,575,006	5.52	1,972	19,668	43,833	76,469	2,072,832	4.06
Pair Race													
Hispanic	2,428	11	1,001	2,663	7,473	732,421	8.38	11	1,182	3,425	9,490	934,741	9.18
black	2,198	12	882	2,630	7,049	396,867	7.59	12	969	3,021	8,122	492,356	7.78
white	12,990	24	814	2,148	5,993	1,575,006	19.99	25	901	2,494	7,238	2,072,832	18.68
other	1,002	7	510	1,681	5,296	113,710	5.63	7	545	2,141	7,203	491,093	9.66
white & black	153	59	987	2,142	4,192	67,379	4.35	152	1,055	2,458	6,412	173,605	6.37
White & Hispanic	592	48	703	2,473	8,327	575,039	11.94	48	716	2,572	8,624	623,207	10.89
white & other	462	31	662	1,688	4,895	213,374	8.01	31	702	1,776	5,827	417,778	11.67
black & Hispanic	61	21	958	2,920	8,065	101,835	4.70	21	1,273	7,853	18,520	124,156	3.49
black & other	73	6	489	1,009	4,253	64,185	5.83	6	558	1,370	4,619	151,882	7.51
Hispanic & other	79	85	554	1,470	4,319	61,728	5.07	85	580	1,686	5,420	76,950	5.17
Pair Gender													
Male, Male	4,366	11	768	2,025	5,466	351,344	6.68	11	845	2,416	6,666	518,490	8.18
Female, Female	4,432	19	796	2,139	5,601	396,867	6.51	19	866	2,402	6,487	466,066	8.48
Male, Female	11,240	6	846	2,347	7,030	1,575,006	18.24	6	945	2,755	8,703	2,072,832	16.64
Household Size													
Two	5,431	31	578	1,601	6,222	157,780	4.64	31	621	1,748	7,844	491,093	6.47
Three	5,176	11	997	2,620	5,903	1,575,006	34.52	11	1,120	3,146	7,178	2,072,832	30.82
Four or more	9,431	6	927	2,360	6,493	1,110,643	10.55	6	1,071	2,798	7,947	1,246,404	10.49

Table L.2 2002 NSDUH Respondent Pair–Level Weight Summary Statistics (continued)

Domain	*n*	Before res.pr.nr[1] (SDUWT*PR02WT10*PR02WT11)						After res.pr.nr[1] (SDUWT*PR02WT10*...*PR02WT12)					
		Min	Q1[2]	Med	Q3[2]	Max	UWE[3]	Min	Q1[2]	Med	Q3[2]	Max	UWE[3]
Census Region													
Northeast	3,982	7	787	2,055	6,181	396,867	5.75	7	862	2,502	7,824	466,066	7.18
South	6,023	6	1,055	2,838	7,500	1,575,006	20.28	6	1,172	3,242	8,757	2,072,832	19.35
Midwest	5,691	31	736	1,831	4,726	1,107,367	19.42	31	808	2,102	5,757	1,207,626	18.81
West	4,342	22	705	2,201	6,813	732,421	8.98	22	784	2,539	8,296	934,741	10.06
Quarter													
Quarter1	4,902	7	794	2,317	7,049	783,080	7.21	7	874	2,734	8,430	903,664	8.43
Quarter2	4,813	6	883	2,277	6,175	1,575,006	19.39	6	972	2,635	7,379	2,072,832	18.63
Quarter3	5,530	21	778	2,089	5,732	732,421	12.29	21	876	2,454	7,107	1,144,702	12.85
Quarter4	4,793	12	818	2,273	6,254	1,572,402	21.56	12	900	2,624	7,758	2,006,610	18.64
% Hispanic in Segment													
50-100%	1,187	11	1,172	3,343	9,656	732,421	10.30	11	1,359	3,976	11,021	934,741	10.14
10-50%	3,277	17	1,055	2,958	8,008	390,814	5.24	17	1,254	3,799	10,482	492,356	6.41
<10%	15,574	6	764	2,053	5,735	1,575,006	19.09	6	827	2,356	6,809	2,072,832	18.46
% Black in Segment													
50-100%	1,554	6	870	2,466	7,234	396,867	7.30	6	942	2,801	8,339	466,066	7.50
10-50%	2,821	11	938	2,574	7,451	1,110,643	11.45	11	1,074	3,116	9,111	1,246,404	10.15
<10%	15,663	7	796	2,148	5,974	1,575,006	17.21	7	879	2,506	7,317	2,072,832	16.60
% Owner-Occupied DUs in Segment													
50-100%	15,169	6	916	2,463	6,726	1,575,006	16.39	6	1,021	2,846	8,252	2,072,832	15.70
10-50%	3,742	12	780	2,051	6,022	261,553	6.24	12	877	2,438	7,556	419,424	7.27
<10%	1,127	11	325	742	2,038	120,616	7.16	11	362	820	2,395	113,946	7.67
Combined Median Rent/Housing Value													
1st Quintile	3,772	22	685	1,707	4,792	1,110,643	21.36	22	741	1,971	5,684	1,246,404	18.03
2nd Quintile	4,114	32	696	1,907	5,454	1,572,402	22.88	32	741	2,126	6,314	2,006,610	22.72
3rd Quintile	4,278	9	863	2,246	6,149	1,575,006	21.48	9	972	2,644	7,580	2,072,832	20.00
4th Quintile	3,892	7	1,044	2,867	8,046	280,277	4.90	7	1,180	3,394	9,797	491,093	6.67
5th Quintile	3,982	6	922	2,592	7,260	1,107,367	11.76	6	1,037	3,030	9,142	1,207,626	11.60
Population Density													
Large MSA	6,978	6	1,235	3,238	8,933	1,575,006	10.45	6	1,385	3,868	11,307	2,072,832	10.47
Medium-Small MSA	7,506	7	749	1,990	5,460	1,572,402	14.99	7	822	2,330	6,529	2,006,610	14.98
Non-MSA, Urban	2,450	14	595	1,675	4,391	110,712	5.24	14	649	1,862	5,143	199,783	6.67
Non-MSA, Rural	3,104	22	544	1,551	4,547	1,110,643	37.57	22	594	1,734	5,208	1,246,404	35.18
Group Quarters													
Group	488	69	481	866	1,761	25,997	3.49	75	517	875	1,950	36,685	4.03
Nongroup	19,550	6	836	2,294	6,404	1,575,006	15.08	6	934	2,672	7,848	2,072,832	14.51

[1] This step used demographic variables from screener data for all selected person pairs; Res = respondent, PR = pair, NR = nonresponse adjustment.
[2] Q1 and Q3 refer to the first and third quartile of the weight distribution.
[3] Unequal weighting effect defined as $1+[(n-1)/n]*CV^2$, where CV=coefficient of variation of weights.

Table L.3 2002 NSDUH Respondent Pair–Level Weight Summary Statistics

Domain	*n*	Before res.pr.ps[1] (SDUWT*PR02WT10*...*PR02WT12)						After res.pr.ps[1] (SDUWT*PR02WT10*...*PR02WT13)						Final Weight: After res.pr.ev[1] (SDUWT*PR02WT10*...*PR02WT14)					
		Min	Q1[2]	Med	Q3[2]	Max	UWE[3]	Min	Q1[2]	Med	Q3[2]	Max	UWE[3]	Min	Q1[2]	Med	Q3[2]	Max	UWE[3]
Total	20,038	6	905	2,601	7,661	2,072,832	14.73	2	853	2,546	7,707	1,506,417	12.77	2	830	2,531	7,735	1,415,150	12.52
Pair Age Group																			
12-17, 12-17	4,192	6	478	1,066	2,328	40,915	2.67	2	435	1,023	2,301	34,422	2.70	2	424	1,010	2,302	36,209	2.75
12-17, 18-25	2,742	22	771	1,641	3,242	39,256	2.48	20	732	1,598	3,334	30,143	2.55	16	716	1,588	3,308	25,984	2.53
12-17, 26-34	694	140	2,131	4,165	8,327	165,945	3.63	114	1,950	4,137	8,177	136,943	3.33	99	1,813	4,004	8,207	148,844	3.51
12-17, 35-49	3,121	104	2,717	5,728	12,147	178,472	2.82	49	2,677	5,909	12,628	165,875	2.73	46	2,580	5,886	12,670	162,334	2.73
12-17, 50+	377	441	5,162	12,939	27,930	160,730	2.39	200	4,997	12,433	27,746	156,924	2.46	161	4,759	12,059	28,622	163,388	2.51
18-25, 18-25	4,419	27	511	1,120	2,844	75,219	3.41	22	468	1,066	2,968	71,872	3.31	18	442	1,051	3,000	76,584	3.29
18-25, 26-34	806	118	1,640	3,760	8,622	176,626	4.31	75	1,544	3,636	8,660	174,387	4.62	64	1,456	3,537	8,750	171,240	4.71
18-25, 35-49	1,042	232	3,434	7,417	18,358	246,940	3.00	138	3,245	7,117	19,119	198,294	3.03	112	3,196	7,136	19,560	179,180	2.95
18-25, 50+	418	512	7,825	15,159	33,272	170,523	2.27	479	6,821	14,956	31,269	208,437	2.45	444	6,704	14,669	31,226	225,948	2.48
26-34, 26-34	559	354	3,668	7,801	14,208	623,207	7.64	246	3,448	8,057	15,600	847,804	8.53	199	3,277	7,753	15,427	847,955	8.51
26-34, 35-49	346	826	5,389	10,412	21,022	934,741	7.90	480	4,361	9,511	18,092	682,999	7.54	463	4,357	9,165	18,647	697,165	7.64
26-34, 50+	123	519	11,478	33,455	72,222	372,961	2.48	329	11,285	36,461	76,730	424,141	2.59	257	10,001	40,625	71,052	428,951	2.58
35-49, 35-49	543	489	7,859	16,430	29,374	2,006,610	12.48	263	6,633	16,210	29,461	1,506,417	10.86	213	6,788	16,309	30,139	1,415,150	10.81
35-49, 50+	210	718	11,115	32,637	73,564	1,207,626	3.94	624	13,423	32,522	72,358	1,214,670	3.80	606	13,618	33,749	73,820	1,039,511	3.38
50+, 50+	446	1,972	19,878	44,347	77,747	2,072,832	4.02	993	19,859	44,221	81,671	1,267,844	2.68	804	19,946	45,667	83,774	1,273,001	2.64
Pair Race																			
Hispanic	2,464	11	1,180	3,441	9,543	934,741	9.08	10	1,143	3,342	9,613	682,999	8.11	8	1,131	3,311	9,775	697,165	8.13
black	2,199	6	971	3,040	8,288	492,356	7.73	4	881	2,859	8,157	596,764	8.96	4	850	2,825	8,196	625,851	9.45
white	12,878	28	896	2,487	7,272	2,072,832	17.68	22	872	2,481	7,391	1,506,417	14.77	18	853	2,469	7,400	1,415,150	14.33
other	922	9	590	2,326	7,743	491,093	9.49	6	531	2,265	8,603	528,936	9.65	5	512	2,270	8,883	451,107	8.77
white & black	168	106	965	2,208	5,179	112,958	5.77	77	952	2,686	5,975	119,352	5.24	74	943	2,754	6,519	121,879	5.07
white & Hispanic	569	25	692	2,337	7,607	623,207	13.12	27	692	2,436	8,665	847,804	15.87	24	643	2,410	8,579	847,955	15.90
white & other	548	7	724	1,904	6,590	1,246,404	28.80	2	341	1,108	3,987	580,015	22.89	2	321	1,051	3,708	606,528	24.55
black & Hispanic	74	21	959	5,109	12,817	173,605	4.87	10	612	4,620	11,667	165,875	4.90	10	658	4,911	12,123	148,152	4.58
black & other	103	12	746	1,736	6,029	79,033	4.16	5	551	1,868	7,652	79,156	4.18	5	496	1,908	7,318	82,127	4.30
Hispanic & other	113	77	623	1,812	5,420	76,950	4.58	36	345	1,083	3,855	101,849	6.95	32	337	983	3,799	105,958	7.31
Pair Gender																			
Male, Male	4,366	11	840	2,404	6,629	518,490	8.27	10	786	2,360	6,857	440,867	7.99	9	766	2,342	6,868	427,670	8.00
Female, Female	4,430	19	872	2,407	6,505	466,066	8.48	20	778	2,365	6,596	570,396	8.72	16	755	2,340	6,638	625,851	8.78
Male, Female	11,242	6	944	2,755	8,704	2,072,832	16.63	2	908	2,710	8,650	1,506,417	14.00	2	885	2,686	8,627	1,415,150	13.65
Household Size																			
Two	5,431	31	621	1,748	7,844	491,093	6.47	18	547	1,656	7,459	528,936	6.77	16	512	1,581	7,350	451,107	6.67
Three	5,176	11	1,120	3,146	7,178	2,072,832	30.82	5	1,055	3,139	7,455	1,506,417	23.13	5	1,053	3,126	7,579	1,415,150	21.22
Four or more	9,431	6	1,071	2,798	7,947	1,246,404	10.49	2	1,019	2,794	7,992	1,093,932	10.41	2	997	2,779	8,003	1,164,378	10.96

Table L.3 2002 NSDUH Respondent Pair–Level Weight Summary Statistics (continued)

Domain	*n*	Before res.pr.ps[1] (SDUWT*PR02WT10*...*PR02WT12)						After res.pr.ps[1] (SDUWT*PR02WT10*...*PR02WT13)						Final Weight: After res.pr.ev[1] (SDUWT*PR02WT10*...*PR02WT14)					
		Min	Q1[2]	Med	Q3[2]	Max	UWE[3]	Min	Q1[2]	Med	Q3[2]	Max	UWE[3]	Min	Q1[2]	Med	Q3[2]	Max	UWE[3]
Census Region																			
Northeast	3,982	7	862	2,502	7,824	466,066	7.18	2	791	2,474	7,791	570,396	7.90	2	777	2,412	7,781	625,851	7.69
South	6,023	6	1,172	3,242	8,757	2,072,832	19.35	4	1,079	3,213	9,066	1,506,417	13.53	4	1,029	3,141	8,928	1,415,150	13.62
Midwest	5,691	31	808	2,102	5,757	1,207,626	18.81	18	790	2,138	5,605	1,214,670	19.74	16	786	2,148	5,644	1,042,479	18.38
West	4,342	22	784	2,539	8,296	934,741	10.06	17	735	2,457	8,539	847,804	9.93	13	710	2,415	8,656	847,955	9.87
Quarter																			
Quarter1	4,902	7	874	2,734	8,430	903,664	8.43	2	808	2,627	8,316	975,487	9.20	2	796	2,636	8,378	991,810	9.03
Quarter2	4,813	6	972	2,635	7,379	2,072,832	18.63	4	940	2,711	7,862	1,267,844	10.98	4	912	2,684	7,869	1,273,001	10.94
Quarter3	5,530	21	876	2,454	7,107	1,144,702	12.85	10	809	2,354	6,986	1,175,490	13.57	10	786	2,304	6,997	1,042,479	13.26
Quarter4	4,793	12	900	2,624	7,758	2,006,610	18.64	5	863	2,596	7,874	1,506,417	17.17	5	841	2,560	7,868	1,415,150	16.69
% Hispanic in Segment																			
50-100%	1,187	11	1,359	3,976	11,021	934,741	10.14	10	1,364	4,025	11,152	847,804	9.72	8	1,345	4,029	11,062	847,955	9.82
10-50%	3,277	17	1,254	3,799	10,482	492,356	6.41	12	1,173	3,771	10,761	596,764	6.51	11	1,171	3,813	10,769	623,397	6.49
<10%	15,574	6	827	2,356	6,809	2,072,832	18.46	2	772	2,296	6,825	1,506,417	15.40	2	753	2,260	6,791	1,415,150	14.99
% Black in Segment																			
50-100%	1,554	6	942	2,801	8,339	466,066	7.50	4	854	2,691	8,291	570,396	8.44	4	821	2,670	8,242	625,851	8.97
10-50%	2,821	11	1,074	3,116	9,111	1,246,404	10.15	9	1,003	3,161	9,572	580,015	7.21	8	949	3,108	9,426	606,528	7.47
<10%	15,663	7	879	2,506	7,317	2,072,832	16.60	2	825	2,453	7,365	1,506,417	14.56	2	814	2,438	7,372	1,415,150	14.10
% Owner-Occupied DUs in Segment																			
50-100%	15,169	6	1,021	2,846	8,252	2,072,832	15.70	2	964	2,792	8,220	1,506,417	13.44	2	941	2,761	8,254	1,415,150	13.16
10-50%	3,742	12	877	2,438	7,556	419,424	7.27	5	807	2,427	7,534	352,695	7.14	5	787	2,410	7,623	340,812	7.03
<10%	1,127	11	362	820	2,395	113,946	7.67	10	314	728	2,491	149,720	8.24	8	282	654	2,431	169,351	9.10
Combined Median Rent/Housing Value																			
1st Quintile	3,772	22	741	1,971	5,684	1,246,404	18.03	17	702	1,954	5,764	975,487	14.37	13	695	1,915	5,764	991,810	14.85
2nd Quintile	4,114	32	741	2,126	6,314	2,006,610	22.72	20	675	2,003	6,234	1,506,417	17.59	16	656	1,955	6,258	1,415,150	17.03
3rd Quintile	4,278	9	972	2,644	7,580	2,072,832	20.00	6	931	2,648	7,956	1,267,844	15.22	5	891	2,621	7,857	1,273,001	15.75
4th Quintile	3,892	7	1,180	3,394	9,797	491,093	6.67	2	1,139	3,401	9,994	442,629	6.71	2	1,132	3,430	10,054	464,286	6.56
5th Quintile	3,982	6	1,037	3,030	9,142	1,207,626	11.60	4	941	2,939	9,206	1,214,670	12.38	4	926	2,957	9,277	1,042,479	11.28
Population Density																			
Large MSA	6,978	6	1,385	3,868	11,307	2,072,832	10.47	4	1,363	3,948	11,601	1,267,844	8.55	4	1,346	3,952	11,497	1,273,001	8.62
Medium-Small MSA	7,506	7	822	2,330	6,529	2,006,610	14.98	2	756	2,247	6,632	1,506,417	12.16	2	720	2,216	6,581	1,415,150	11.75
Non-MSA, Urban	2,450	14	649	1,862	5,143	199,783	6.67	9	609	1,840	5,387	199,638	6.48	8	587	1,822	5,261	193,210	6.58
Non-MSA, Rural	3,104	22	594	1,734	5,208	1,246,404	35.18	17	556	1,642	5,203	1,214,670	35.71	13	544	1,632	5,164	1,164,378	33.85
Group Quarters																			
Group	488	75	517	875	1,950	36,685	4.03	37	395	710	1,805	38,767	5.21	38	396	677	1,823	38,992	4.90
Nongroup	19,550	6	934	2,672	7,848	2,072,832	14.51	2	885	2,633	7,906	1,506,417	12.58	2	862	2,606	7,926	1,415,150	12.33

Table L.3 2002 NSDUH Respondent Pair-Level Weight Summary Statistics (continued)

Domain	*n*	Before res.pr.ps[1] (SDUWT*PR02WT10*...*PR02WT12)						After res.pr.ps[1] (SDUWT*PR02WT10*...*PR02WT13)						Final Weight: After res.pr.ev[1] (SDUWT*PR02WT10*...*PR02WT14)					
		Min	Q1[2]	Med	Q3[2]	Max	UWE[3]	Min	Q1[2]	Med	Q3[2]	Max	UWE[3]	Min	Q1[2]	Med	Q3[2]	Max	UWE[3]
Pair Relationship[4]																			
Parent-child (12-14)	2,133	140	2,495	5,269	11,265	178,472	3.09	102	2,565	5,540	11,841	165,875	3.01	82	2,487	5,482	11,818	163,388	3.03
Parent-child (12-17)	3,864	140	2,610	5,636	12,062	178,472	3.01	102	2,697	5,946	12,815	165,875	2.92	82	2,622	5,939	12,839	163,388	2.94
Parent-child (12-20)	4,545	140	2,787	5,939	13,163	246,940	3.16	102	2,853	6,416	14,038	208,437	3.15	82	2,803	6,371	14,104	225,948	3.15
Sibling (12-14) - sibling (15-17)	2,464	7	477	1,049	2,280	24,109	2.49	2	451	1,052	2,347	23,196	2.51	2	440	1,036	2,321	23,685	2.54
Sibling (12-17) - sibling (18-25)	2,425	36	794	1,645	3,211	39,256	2.47	27	730	1,591	3,227	30,143	2.53	23	714	1,574	3,224	25,984	2.52
Spouse-spouse	4,000	27	791	2,698	12,587	2,072,832	18.79	22	804	2,735	12,870	1,506,417	14.72	18	785	2,715	12,675	1,415,150	14.21
Spouse-spouse with children (under 18)	2,020	27	787	2,383	8,998	2,006,610	31.96	26	919	2,725	10,262	1,506,417	23.65	22	913	2,758	10,280	1,415,150	23.30

[1] This step used demographic variables from questionnaire data for all selected person pairs; Res = respondent, PR = pair, PS = poststratification adjustment, EV = extreme value adjustment.
[2] Q1 and Q3 refer to the first and third quartile of the weight distribution.
[3] Unequal weighting effect defined as $1+[(n-1)/n]*CV^2$, where CV=coefficient of variation of weights.
[4] Parent-child (15-17) was not included here since extreme values were not controlled with this domain. Spouse-spouse pair relationships also included partner-partner relationships.

Appendix M

Hot-Deck Method of Imputation

Appendix M
Hot-Deck Method of Imputation

M.1 Introduction

Typically, with the hot-deck method of imputation, missing responses for a particular variable (called the "base variable" in this appendix) are replaced by values from similar respondents with respect to a number of covariates (called "auxiliary variables" in this appendix). If "similarity" is defined in terms of a single predicted value from a model, these covariates can be represented by that value. The respondent with the missing value for the base variable is called the "recipient," and the respondent from whom values are borrowed to replace the missing value is called the "donor."

Three different methods of hot-deck imputations are discussed in this document, though only two were used in the 2002 National Survey on Drug Use and Health (NSDUH)[16]: unweighted sequential hot deck, unweighted random nearest neighbor hot deck (NNHD), and weighted sequential hot deck. The first method, the unweighted sequential hot deck, was the exclusive method of hot-deck imputation used for the 1991 to 1998 surveys and the paper-and-pencil interviewing (PAPI) sample of the 1999 survey. This method was used for all demographic variables in the 1999 survey, but no other variables. In the 2000 NSDUH, the unweighted sequential hot deck method was only used for education and employment status, and was not used at all in 2001 or 2002 surveys. However, it remains in this appendix for historical purposes and for the sake of comparison with the other two methods. In a similar manner to the 1999 (computer-assisted interviewing [CAI] sample of the survey), 2000, and 2001 surveys, the 2002 NSDUH primarily used the second hot-deck method listed, the unweighted random NNHD. The third hot-deck method, weighted sequential hot deck, incorporated the sampling weights associated with each respondent. Starting in the 2002 NSDUH, the immigrant variable imputations described in Chapter 5 utilized the weighted sequential hot-deck method. For more information on weighted sequential hot-deck see Cox (1980, pp. 721-725) and Iannacchione (1982).

A step that is common to all hot-deck methods is the formation of imputation classes, which is discussed in Section A.2. This is followed by a general description of the three hot-deck methods Sections A.3-A.5. With each type of hot-deck imputation, the identities of the donors are generally tracked. For more information on the general hot-deck method of item imputation, see Little and Rubin (1987, pp. 62-67).

M.2 Formation of Imputation Classes

When there was a strong logical association between the base variable and certain auxiliary variables, the dataset was partitioned by the auxiliary variables and imputation

[16] This report presents information from the 2002 National Survey on Drug Use and Health (NSDUH), an annual survey of the civilian, noninstitutionalized population of the United States aged 12 years old or older. Prior to 2002, the survey was called the National Household Survey on Drug Abuse (NHSDA).

procedures were implemented independently within classes defined by the cross of the auxiliary variables. These classes were defined by logical and likeness constraints, which are described in the main body of this report. Classes defined by the likeness constraints were collapsed if insufficient donors were available, and classes defined by logical constraints were not collapsed, due to the possibility of an inconsistency with pre-existing nonmissing values that would have resulted.

M.3 Unweighted Sequential Hot Deck

In the years that the unweighted sequential hot deck was used, its implementation involved three basic steps. After the imputation classes were formed, the file was appropriately sorted and imputed values assigned, as described in the following sections.

M.3.1 Sorting the File

Within each imputation class, the file was sorted by auxiliary variables relevant to the item being imputed. The sort order of the auxiliary variables was chosen to reflect the degree of importance of the auxiliary variables in their relation to the base variable being imputed (i.e., those auxiliary variables that were better predictors for the item being imputed were used as the first sorting variables). In general, two types of sorting procedures were used in previous NSDUHs to sort the files prior to imputation:

- **Straight Sort.** A set of variables was sorted in ascending order by the first variable specified; then within each level of the first variable, the file was sorted in ascending order by the second variable specified; and so forth. For example:

  1 1 1
  1 1 2
  1 2 1
  1 2 2
  1 3 1
  1 3 2
  2 1 1
  2 1 2
  2 2 1
  2 2 2
  2 3 1
  2 3 2

- **Serpentine Sort.** A set of variables was sorted so that the direction of the sort (ascending or descending) changes each time the value of a variable changes. For example:

  1 1 1
  1 1 2
  1 2 2
  1 2 1
  1 3 1

1	3	2
2	3	2
2	3	1
2	2	1
2	2	2
2	1	2
2	1	1

The serpentine sort has the advantage of minimizing the change in the entire set of auxiliary variables every time any one of the variables changes its value.

M.3.2 Replacing Missing Values

The file was sorted and then read sequentially. Each time an item respondent was encountered (i.e., the base variable was nonmissing), the base variable response was stored, updating the donor response, and any subsequent nonrespondent that was encountered received the stored donor response creating the statistically imputed response. A starting value was needed if an item nonrespondent was the first record on a sorted file. Typically, the response from the first respondent on the sorted file was used as the starting value. Due to the fact that the file was sorted by relevant auxiliary variables, the preceding item respondent (donor) closely matched the neighboring item nonrespondent (recipient) with respect to the auxiliary variables.

M.3.3 Potential Problem

With the unweighted sequential hot-deck imputation procedure, for any particular item being imputed, there was the risk of several nonrespondents appearing next to one another on the sorted file. To detect this problem in the NSDUH, the imputation donor was identified for every item being imputed. Then, when frequencies by imputation donor were examined, the problem was detected if several nonrespondents were aligned next to one another in the sort. When this problem occurred, sort variables were added or eliminated, or the order of the variables was rearranged.

M.4 Unweighted Random Nearest Neighbor Hot Deck

As with the unweighted sequential hot deck, the unweighted random NNHD was implemented in three steps. After the imputation classes were formed, a neighborhood of potential donors was created, from which imputed values were assigned, as described in the following sections.

M.4.1 Creating a Neighborhood of Potential Donors

First, a metric was defined to measure the distance between units, based on the values of the covariates. Then a neighborhood was created of potential donors "close to" the recipient based on that metric. For example, the distance between the values of the recipient and potential donors for each of the auxiliary variables were calculated, then the donors for the neighborhood were chosen such that the maximum of these distances was less than a certain value, referred to as "delta." This neighborhood was restricted, using the imputation classes defined above, so that

the potential donors' values of the base variable were consistent with the recipient's preexisting nonmissing values of related variables. In the NSDUH, the values of the auxiliary variables were represented by a predicted mean from a model, so that the distance metric was a univariate Euclidean distance between the predicted mean of the recipient and the potential donors. The distance was relative when dividing this value by the predicted mean of the recipient, resulting in delta as a percentage.

M.4.2 Randomly Selecting a Donor for the Recipient from the Neighborhood of Donors

From the neighborhood of donors created in the previous step, a single donor was randomly selected. The base variable values for this single donor replaced those of the recipient. The selection was conducted as a simple random sample because weights were incorporated in determining the neighborhood mean, which was the predicted mean. Alternatively, a weighted selection could have been employed if weights had not been used to determine the neighborhood mean.

M.5 Weighted Sequential Hot Deck

The steps taken to impute missing values in the weighted sequential hot deck were equivalent to those of the unweighted sequential hot deck. The details on the final imputation, however, differed with the incorporation of sampling weights. The first step, as always, was the formation of imputation classes. Afterwards, two additional steps, as described below, were implemented.

M.5.1 Sorting the File

Within each imputation class, the file was sorted by auxiliary variables relevant to the item being imputed. The sort order of the auxiliary variables was chosen to reflect the degree of importance of the auxiliary variables in their relation to the base variable being imputed (i.e., those auxiliary variables that were better predictors for the item being imputed were used as the first sorting variables). In general, two types of sorting procedures were used in previous NSDUHs to sort the files prior to imputation: straight sort and serpentine sort. Both of these methods are described in detail in Section A.2.2.

M.5.2 Replacing Missing Values

The procedure used in the 2002 NSDUH followed directly from Cox (1980). Specifically, once the imputation classes are formed, the data is divided into two data sets: one for respondent and one for nonrespondents. Scaled weights $v(j)$ are then derived for all nonrespondents using the following formula:

$$v(j)=w(j)s(+)/w(+);\ j=1, 2, \ldots n$$

where n is the number of nonrespondents, $w(j)$ is the sample weight for the j^{th} nonrespondent, $w(+)$ is the sum of the sample weights for the all nonrespondents, and $s(+)$ is the sum of the sample weights for all the respondents (Cox, 1980). The respondent data file is partitioned into zones of width $v(j)$, where the imputed value for the j^{th} nonrespondent is selected from a respondent in the corresponding zone of the respondent data file.

This selection algorithm is an adaptation of Chromy's (1979) sequential sample selection method, which could be implemented using the Chromy-Williams sample selection software (Williams and Chromy, 1980). Furthermore, Iannacchione (1982) revised the Chromy-Williams sample selection software, so that each step of the weighted sequential hot deck is executed in one macro run.

M.5.3 Benefits of Weighted Sequential Hot-Deck

With the unweighted sequential hot-deck imputation procedure, for any particular item being imputed, there is the risk of several nonrespondents appearing next to one another on the sorted file. An imputed value could still be found for those cases, since the algorithm would select the previous respondent in the file; however, some modifications are required in the sorting procedure to prevent a single respondent from being the donor for several nonrespondents (see Section A.3.3). With the weighted sequential hot-deck method, on the other hand, this problem does not occur because the weighted hot-deck controls the number of times a donor can be selected. In addition, the weighted hot-deck allows each respondent the chance to be a donor since a respondent is selected within each $v(j)$.

The most important benefit of the weighted sequential hot-deck method, however, is the elimination of bias in the estimates of means and totals. This type of bias is particularly present when the response rate is low or the covariates explain only a small amount of variation in the specified variable. In addition, many surveys sample subpopulations at different rates and using the sample weights allows, in expectation, the imputed data for the nonrespondents to have the same mean (for the specified variables) as the respondents. In other words, the weighted hot-deck preserves the respondent's weighted distribution in the imputed data (Cox, 1980).

Appendix N

Univariate and Multivariate Predictive Mean Neighborhood Imputation Methods

Appendix N

Univariate and Multivariate Predictive Mean Neighborhood Imputation Methods

N.1 Introduction

The 2002 National Survey on Drug Use and Health (NSDUH)[17] used a predictive mean neighborhood (PMN) method for imputing missing values. This method was implemented in the past several surveys. Starting with the 1999 survey, this PMN method was a new approach, which was developed for the imputation of missing values in the computer-assisted interviewing (CAI) sample. This approach has been used since the 1999 NSDUH[18] and can be applied to one variable at a time or to several variables simultaneously. As described in this appendix, it incorporates predictive means from models and the assignment of imputed values using neighborhoods determined by those predictive means.

N.2 Overview

N.2.1 Predictive Mean Neighborhoods:Derived from Combining Nearest Neighbor Hot Deck and Predictive Mean Matching

The PMN method is a combination of two commonly used imputation methods: a non-model-based hot deck (nearest neighbor), and a modification of the model-assisted predictive mean matching (PMM) method of Rubin (1986). PMN enhances the PMM method in that it can be applied to both discrete and continuous variables either individually or jointly. PMN also enhances the nearest neighbor hot-deck (NNHD) method in that the distance function used to find neighbors is no longer ad hoc.

A commonly used imputation method is a random NNHD (Little & Rubin, 1987, p. 65). With this method, donors and recipients are distinguished by the completeness of their records with regard to the variable(s) of interest (the donor has complete data, the recipient does not). A donor set deemed close to the recipient with respect to a number of covariates is used to select a donor at random. For the NSDUH, the set of covariates typically included demographic variables, as well as some other nonmissing drug use variables. In the case of the NSDUH, to further ensure that a donor matched the recipient as closely as possible, discrete variables (or discrete categories of continuous variables) strongly correlated with drug use, such as age categories, were often used to restrict the set of donors. Furthermore, other restrictions involving outcome variables were imposed on the neighborhood.

[17] This report presents information from the 2002 National Survey on Drug Use and Health (NSDUH), an annual survey of the civilian, noninstitutionalized population of the United States aged 12 years old or older. Prior to 2002, the survey was called the National Household Survey on Drug Abuse (NHSDA).

[18] In the surveys after the 1999 one, only a CAI sample was selected.

Note that in NNHD, unlike sequential hot deck, a distance function is used to define closeness between the recipient and a donor. So, there is less of a problem of sparseness of the donor class, but the distance function involving categorical or nominal variables is typically ad hoc and often hard to justify.

The PMM method is only applicable to continuous outcome variables. With this method, a distance function is used to determine distances between the predictive mean for the recipient, obtained under a model, and the response variable outcomes for candidate donors. The respondent with the smallest distance is chosen as the donor. Unlike the NNHD, the donor is not randomly selected from a neighborhood. The advantages of PMM include the following:

- Model bias in the predictive mean can be minimized by using suitable covariates.
- The PMM method is not a pure model-based method because the predictive mean is only used to assist in finding a donor. Hence, like NNHD, it has the flexibility of imposing certain constraints on the set of donors.

However, the choice of donor is nonrandom. This nonrandomness leads to bias in the estimators of means and totals. It also tends to make the distribution of outcome values skewed to the center. Furthermore, as mentioned earlier, the PMM method is not applicable to discrete variables, because the distance function between the recipient's predictive mean (which takes continuous values) and the donor's outcome value (which takes discrete values) is not well defined.

N.2.2 Univariate and Multivariate Applications of Predictive Mean Neighborhoods

PMN method is easily applicable to problems of both univariate and multivariate imputations. The need for univariate imputation arises when the value of a single continuous variable, such as age at first use of marijuana, or a single dichotomous discrete variable, such as lifetime use of marijuana, is missing for a respondent. On the other hand, the need for multivariate imputation arises when values of two or more variables are missing for a single respondent. The case of a single polytomous variable, such as marijuana recency of use with missing values, can also be viewed as a multivariate imputation problem.

The standard approach to multivariate modeling, with a given set of outcome variables (including both discrete and continuous), is likely to be tedious in practice because of the computational problems due to the volume of model parameters, and the difficulty in specifying a suitable covariance structure. Following Little and Rubin's (1987) proposal of a joint model for discrete and continuous variables, and its implementation by Schafer (1997), it is possible to fit a pure multivariate model for multivariate imputation, but it would require making distributional assumptions. Moreover, none of the existing solutions takes the survey design into account because of the obvious problem of specifying the probability distribution underlying survey data. However, in the application of the multivariate predictive mean neighborhood (MPMN) imputation to the 1999-2002 surveys, a multivariate model was fitted by a series of univariate parametric models (including the polytomous case), such that variables modeled earlier in the hierarchy had a chance to be included in the covariate set for subsequent models in the hierarchy. In the multivariate modeling with MPMN, the innovative idea is to express the likelihood in the

superpopulation model as a product of marginal and conditional likelihoods, which then allows for use of univariate techniques for fitting multivariate (but conditional) predictive means.

If it turns out that a donor set for MPMN is sparse, the univariate predictive mean neighborhood (UPMN) procedure can be used as an alternative. Assuming that the donor set (i.e., the set of complete records in a small neighborhood of the recipient with respect to all the elements of the predictive mean) is not sparse, having a single record to fill all the missing values in an incomplete record is desirable because this method preserves the relationships among the variables of interest. Moreover, if the predictive mean vector includes both missing and nonmissing variables (this could easily happen when models are fitted in a univariate manner under a hierarchy), it is also ensured that the predictive mean vector for the donor record is not only close to the recipient with respect to missing variables, but also with respect to the nonmissing ones. Although the nonmissing values would not be replaced by the corresponding values from the donor, some degree of correlation between missing and nonmissing variables is expected to be preserved because of the closeness between the donor and the recipient. This is due to the fact that the predictive mean vector consists of conditional means (the drug use covariates in the conditioning set appear earlier on in the hierarchy); therefore, being close to the conditional means should help in preserving the correlation among outcome variables on the recipient record.

N.3 Outline and Description of Method

The procedure for implementing UPMN and MPMN in the 2002 NSDUH entailed six steps. Steps 2 through 5, and sometimes Step 6, were cycled through each of the drugs and drug use measures in the order determined by Step 1. Steps 4 and 5 (Steps 4 to 6 when applicable) could have been considered a variant of a random NNHD.

N.3.1 Step 1: Definition of Hierarchy

The first step was to determine the order in which variables were modeled, so that variables early in the hierarchy could have been used for modeling the conditional predictive mean (i.e., they have the potential to have been part of the set of covariates for variables later in the hierarchy). Note that usually not all variables in the hierarchy were missing for a particular incomplete record. Nevertheless, models were developed for all the variables in a univariate fashion for reasons mentioned earlier. For example, in the drug modules in the 2002 NSDUH, different drugs needed to have been modeled, with different measures of drug use for each drug. It was therefore necessary to determine the order in which the combination of drugs and drug use measures would have been handled. Using the sequence of variables determined by this step, the procedure involved cycling through Steps 2 through 5, and sometimes Step 6. In the application of the PMN to the NSDUH, the order of imputation for drugs was determined by considering such factors as the level of stigma associated with the drugs, the level of "missingness" in the data (see Appendix G), and the degree to which one set of drugs could have been used as predictors for other drugs. The order of drugs was given by cigarettes, smokeless tobacco, cigars, pipes, alcohol, inhalants, marijuana, hallucinogens, pain relievers, tranquilizers, stimulants, sedatives, cocaine, crack, and heroin. The order of drug use measures imputed was determined based on the natural hierarchy of the variables: lifetime usage, recency of use, frequency of use in the past 12 months, frequency of use in the past 30 days, and age of first use.

For each variable, Steps 2 through 5 were followed for the NSDUH.

N.3.2 Step 2: Setup for Model Building and Hot-Deck Assignment

For each model that was fitted, two groups were created: complete data respondents and incomplete data respondents (item respondents and item nonrespondents, respectively). Complete data respondents had complete data across the variables of interest, and incomplete data respondents encompassed the remainder of respondents. If the final assignment was multivariate, complete data respondents must have had complete data across all the variables in the multivariate response vector. Models were constructed using complete data respondents only.

N.3.3 Step 3: Sequential Hierarchical Modeling

The model was built using the complete data respondents only with weights adjusted for item nonresponse. For the drug modules in the 2002 NSDUH, lifetime usage indicators were modeled first because all other drug use indicators depended on an indication of lifetime use or nonuse. Once the hierarchy of drugs for lifetime usage was determined, lifetime usage indicators for individual drugs were modeled in a sequential fashion. The sequence used for the remaining combinations of drugs and drug use measures depended on what covariates were desired in the models and what variables were considered part of a multivariate set.

N.3.4 Step 4: Computation of Predictive Means and Delta Neighborhoods

Once the model was fitted, the predictive means for item respondents and item nonrespondents were calculated using the model coefficients. For models with a multivariate predictive mean vector (such as with a polytomous logit model), a single element out of that vector was chosen, so that each respondent had exactly one predictive mean value.[19] This predictive mean was the matching variable in a random NNHD. It could have come directly from the model, it could have been adjusted to account for the conditioning on the time period, or (if it was the predicted value based on a model with a transformed response variable) it could have been back-transformed to the original units.

For each item nonrespondent, a distance was calculated between the predictive mean of the item nonrespondent and the predictive means of every item respondent. Those item respondents, whose predictive means were "close" (within a predetermined value delta) to the item nonrespondent, were considered as part of the "delta neighborhood" for the item nonrespondent and were potential donors. If the number of item respondents who qualified as donors was greater than some number, say k, only those item respondents with the smallest k distances were eligible donors.

The pool of donors was further restricted to satisfy constraints to make imputed values consistent with the preexisting nonmissing values of the item nonrespondent. An example of this type of constraint, called a "logical constraint," was given by age at first crack use, which must

[19] Alternatively, a provisional MPMN method could have been performed by using the predicted probabilities from the polytomous model. Consequently, the final MPMN would have been built based on probabilities from the polytomous model, as well as predictive means for the other variables in the multivariate set. See Step 6 (Section C.3.6) for a description of the MPMN.

not have been less than age at first cocaine use. Other constraints, called "likeness constraints," were placed on the pool of donors to make the attributes of the neighborhood as close to that of the recipient as possible. For example, for age at first use, the age of the donor and the age of the recipient were restricted to have been the same whenever possible, and the donor and recipient must have come from States with similar usage patterns. A small value of delta could have also been considered as a likeness constraint. Whenever insufficient donors were available to meet the likeness constraints, including the preset small value of delta, the constraints were loosened in priority order according to their perceived importance. As a last resort, if an insufficient number of donors were available to meet the logical constraints given the loosest set of likeness constraints allowable, a donor was found using a sequential hot deck, where matching was done on the predictive mean. (Even though weights would not have been used to determine the donor in the sequential hot deck, "unweighted" is not an accurate characterization of the imputation process because weighting would already have been incorporated in the calculation of the predicted mean.)

If many variables were imputed in a single multivariate imputation, it was advantageous to preserve, as much as possible, correlations between variables in the data. However, the more variables that were included in a multivariate set, the less likely that a neighborhood could have been used for the imputation within a given delta. Even though there were many advantages to using multivariate imputation, one disadvantage, in several instances, was not being able to find a neighborhood within the specified delta.

N.3.5 Step 5: Assignment of Imputed Values Using a Univariate Predictive Mean Neighborhood

Using a simple random draw from the neighborhood developed in Step 4, a donor was chosen for each item nonrespondent. If only one response variable was imputed, the assignment step was a simple replacement of a missing value by the value of the donor. It was possible, however, that a donated quantity was a function of the final imputed value. For example, for 12-month frequency of drug use, because donors and recipients could potentially have had a different maximum possible number of days in the year that they could have used a substance, the observed proportion of total period was donated rather than the observed 12-month frequency, where the "total period" could have ranged up to a year. In the assignment step, the donor's proportion of total period was multiplied by the recipient's maximum possible number of days in the year that he or she could have used the substance.

The assignment step was multivariate if several response variables were associated with a single predictive mean, provided more than one of those response variables was missing. In that case, all of the missing values were imputed using the same donor. If there was more than one response variable associated with a single predictive mean, but not all of them were missing, only the missing values were replaced by those of the donor. The resulting imputed values were

provisional if a multivariate predictive mean vector was needed in a final multivariate imputation; otherwise, these values were final.[20]

The variables requiring imputation were part of a multivariate set if a multivariate predictive mean vector was used to match donors and recipients in a final multivariate imputation. If the variables were part of a multivariate set, then it was necessary to cycle through Steps 2 through 5 for each variable in the set, then proceed to Step 6 after completing Steps 2 through 5 for the last variable in the set. If the variables were not part of a multivariate set, then it was only necessary to go through Steps 2 through 5 once, and proceeding to Step 6 was unnecessary. After the completion of either Step 5 (if a univariate predicted mean was used) or Step 6 (if a multivariate predictive mean vector was used), the next variable in the hierarchy requiring imputation was processed by returning to Step 2.

N.3.6 Step 6: Determination of Multivariate Predictive Mean Neighborhood and Assignment of Imputed Values

With the MPMN method, the neighborhood was defined based on a vector of predictive means rather than from a single predictive mean as in the univariate case. This vector may have encompassed a subvector of predictive means from a single categorical model (as with a polytomous logit model), in addition to scalar predictive means from any number of models with continuous response variables. For each item nonrespondent, a distance was calculated between the elements of this vector of predictive means, where the observed values were missing, and the corresponding elements of the vector for every item respondent. To make all elements of the vector conditional on the same usage status in the full predictive mean vector, predictive means that were calculated on the basis of past year and past month users were furthermore adjusted to account for the probability that a respondent was a past year user or a past month user. For example, in the 2002 NSDUH, the full predictive mean vector for alcohol included the following elements:

1. *recency, past month: P* (past month alcohol user | lifetime alcohol user);
2. recency, past year, not past month: *P* (past year but not past month alcohol user | lifetime alcohol user);
3. *12-month frequency: P* (the respondent used alcohol on a given day in the past year | past year user of alcohol) * *P* (past year user of alcohol | lifetime alcohol user)[21];
4. *30-day frequency: P* (the respondent used alcohol on a given day in the past month | past month user of alcohol) * *P* (past month alcohol user | lifetime alcohol user); and

[20] If the variable was part of a multivariate set upon which the MPMN method was applied, and provisional values were not needed for subsequent models, Steps 4 (creation of delta neighborhood) and 5 could have been skipped.

[21] For the 12-month frequency, 30-day frequency, and 30-day binge frequency, the models were fitted using logits. These logits were converted to probabilities when creating the predictive mean vector. Interpreting the proportion of the year used as a probability of use on a given day in the year assumed that the probability of use on each day in the year was equal. This, of course, was not true. However, the violation of this assumption did not seriously affect the ability to find a reasonable variable to use for finding a neighborhood, and it did allow a predicted mean to be made conditional on what was known.

5. *30-day binge frequency: P* (the respondent was a binge drinker on a given day in the past month | past month user) * *P* (past month alcohol user | lifetime alcohol user).

The subset of elements used to determine a neighborhood for a particular item nonrespondent depended on the missingness pattern of that item nonrespondent.[22] Moreover, if partial information was available on the recency of use, the predictive means was adjusted to account for that knowledge. For example, if a particular item nonrespondent was known as a past year alcohol user and his 12-month frequency was known, the elements above for which differences would have been calculated would be element #1 conditioned on past year use, and elements #4 and #5. That is,

P (Past month alcohol user | Lifetime alcohol user) ÷ P (Past year alcohol user | Lifetime alcohol user),

P (Respondent used alcohol on a given day in the past month | Past month user of alcohol) P (Past month alcohol user | Lifetime alcohol user) ÷ P (Past year alcohol user | Lifetime alcohol user), and*

*P (Respondent was a binge drinker on a given day in the past month | Past month user) * P (Past month alcohol user | Lifetime alcohol user) ÷ P (Past year alcohol user | Lifetime alcohol user).*

A neighborhood that resulted from this vector of distances was constrained by a multivariate preset delta, where the distances associated with each element of the predictive mean vector must each have been less than the preset delta associated with that element. From the donors that remained, a single neighborhood was created out of a vector of differences by converting that vector to a scalar, called the Mahalanobis distance, which is given by

$$(\mu_R - \mu_{NR})^T \Sigma^{-1} (\mu_R - \mu_{NR})$$

where μ_R refers to the predictive mean (sub-)vector for a given item respondent, and μ_{NR} is the predictive mean (sub-)vector for a given item nonrespondent. The matrix Σ is the variance-covariance matrix of the predictive means, calculated using the subvector of predictive means associated with each missingness pattern, using complete data respondents within each age group and (where applicable) State rank group. The Mahalanobis distance was only calculated for those respondents who met the delta constraint. The neighborhood was determined by selecting the *k* smallest Mahalanobis distances within this subset of item respondents for a given item nonrespondent.

For those variables in the response vector that were not missing, only those that were missing were replaced. However, logical constraints must have been placed on the multivariate neighborhood, so that imputed values were consistent with preexisting nonmissing values. For example, if a respondent was missing a 30-day frequency, but his or her nonmissing 12-month

[22] Alternatively, the entire predictive mean vector could have been used to determine the neighborhood, regardless of the missingness pattern. Due to the fact that many respondents in the multivariate set were only missing one item in the set, imputation was accomplished using UPMN, which is computationally much faster.

frequency was 350, a donor could not have had a 30-day frequency smaller than 350 - 335, or 15. If the number of respondents in the univariate subset who met the logical constraints, imposed upon the multivariate neighborhood, was fewer than *k* but greater than 0, all the respondents in the resulting subset were selected for the neighborhood. Finally, if there were no respondents within the univariate subset who met the logical constraints imposed by the multivariate neighborhood, the *k* smallest Mahalanobis distances who met the logical constraints among all candidate donors for a given item nonrespondent were selected for the neighborhood. In addition to the multivariate delta, likeness constraints were used to make the donors in the neighborhood as much like the recipient as possible. These could have been loosened if insufficient donors were available. Finally, as with the univariate neighborhood, an unweighted sequential hot deck was used as a last resort if there were not enough sufficient donors available who met the logical constraints and the loosest set of likeness constraints allowable.

As with the univariate assignments, a donor was randomly drawn from the neighborhood for each item nonrespondent. For most variables, the observed value of interest was donated directly to the recipient. As in the univariate case, however, it was possible for a donated value to have been a function of the final imputed value, rather than the imputed value itself. The 12-month frequency example given in Step 5 applies here as well.

N.4 Comparison of PMN with Other Available Imputation Methods

The PMN methodology addresses all of the shortcomings of the unweighted sequential hot-deck method:

- **Ability to use covariates to determine donors is far greater than in the hot deck.** As with other model-based techniques, using models allows more covariates to be incorporated, including measures of use of other drugs, in a systematic fashion, where weights can be incorporated without difficulty. However, like a hot deck, covariates not explicitly modeled can be used to restrict the set of donors using logical constraints. If there is particular interest in having donors and recipients with similar values of certain covariates, they can be used to restrict the set of donors using likeness constraints even if they are already in the model.
- **Relative importance of covariates is determined by standard estimating equation techniques.** In other words, there are objective criteria based on methodology, such as regression, that quantify the relationship between a given covariate and the response variable, in the presence of other covariates. Thus, the response variable itself is indirectly used to determine donors.
- **Problem of sparse neighborhoods is considerably reduced, which makes it easier to implement restrictions on the donor set.** Because the distance function is defined as a continuous function of the predictive mean, it is possible to find donors arbitrarily close to the recipient. Thus, it is less likely to have the problem of sparse neighborhoods for hot decking. Moreover, having sufficient donors in the neighborhood allows for imposing extra constraints on the donor set, which would be difficult to incorporate directly in the model.

- **Sampling weights are easily incorporated in the models.** The weighted hot deck can be viewed as a special case of PMN.
- **Correlations across response variables are justified by making the imputation multivariate.**
- **Choice of donor can be made random by choosing delta large enough such that the neighborhood is of a size greater than 1.** Under the assumption that the recipient and the candidate donors in the neighborhood have approximately equal means, the random selection allows the case where the error distribution with mean zero can be mimicked. This helps to avoid bias in estimating means and totals, variances of which can be estimated as in two-phase sampling or by suitable resampling methods.

In comparison with other model-based methods, discrete and continuous variables can be handled jointly and relatively easily in MPMN by using the idea of univariate (conditional) modeling in a hierarchical manner. In MPMN, differential weights can be objectively assigned to different elements of the predictive mean vector depending on the variability of predictive means in the dataset via the Mahalanobis distance.

As noted earlier, the PMN method has some similarity with the predictive mean matching method of Rubin (1986) except that, for the donor records, the observed variable value and not the predictive mean is used for computing the distance function. Also, the well-known method of nearest neighbor imputation is similar to PMN, except that the distance function is in terms of the original predictor variables and would often require arbitrary scaling of discrete variables. Moreover, for this method, it is generally hard to objectively decide about the relative weights for different predictor variables.

Appendix O

Rules for Determining Pair Relationships

Appendix O

Rules for Determining Pair Relationships

O.1 Rules for determining matching pairs, in priority order

The following rules are used to determine the roster member in a respondent's household roster that corresponds to the other pair member. In the rules that follow, an "age match" occurs if the questionnaire age of one pair member matches a roster age in the other pair member's roster, and a "gender match" occurs if the questionnaire gender of one of the pair members matches a roster gender in the other pair member's roster. In the following table, if the rules for Pair Member A and Pair Member B in a single row differ, then the count for that row includes the rules as listed, and the rules with Pair Member A and Pair Member B reversed. If the age and/or gender are off when finding these matches, the age and/or gender are defined by the questionnaire age and gender of the selected pair member when determining the pair domain. The rules are listed in priority order in Exhibit O.1, along with the number of pairs to which each rule was applied. Since the 2001 survey, it was technically impossible to identify more than one roster member as the "other pair member selected," resulting in either 0 or 1 MBRSEL for each responding pair. Rules involving situations where more than one MBRSEL existed are therefore not included in this table. Some other conditions which were not manifest in 2002 are also excluded from this table, provided the distribution of counts would have been unaffected by their exclusion from the code.

Exhibit O.1 Rules for Determining Matching Pairs, in Priority Order

Priority	Rule		
	Pair Member A	Pair Member B	Count
1	Age and gender match exactly, exactly one MBRSEL in right place	Age and gender match exactly, exactly one MBRSEL in right place	17,045
2	Age and gender match exactly, exactly one MBRSEL in right place	Age within one, gender matches exactly, exactly one MBRSEL in right place	1,809
3	Age within one, gender matches exactly, exactly one MBRSEL in right place	Age within one, gender matches exactly, exactly one MBRSEL in right place	121
4	Age and gender match exactly, exactly one MBRSEL in right place	Age within two, gender matches exactly, exactly one MBRSEL in right place	220
5	Age within one, gender matches exactly, exactly one MBRSEL in right place	Age within two, gender matches exactly, exactly one MBRSEL in right place	33
6	Age within two, gender matches exactly, exactly one MBRSEL in right place	Age within two, gender matches exactly, exactly one MBRSEL in right place	5

Exhibit O.1 Rules for Determining Matching Pairs, in Priority Order (continued)

Priority	Rule		
	Pair Member A	**Pair Member B**	**Count**
7	Age and gender match exactly, exactly one MBRSEL in right place	Age and gender match exactly, MBRSEL missing for all roster members	380
8	Age within one, gender matches exactly, exactly one MBRSEL in right place	Age and gender match exactly, MBRSEL missing for all roster members	28
9	Age within two, gender matches exactly, exactly one MBRSEL in right place	Age and gender match exactly, MBRSEL missing for all roster members	2
10	Age and gender match exactly, MBRSEL missing for all roster members	Age and gender match exactly, MBRSEL missing for all roster members	37
11	Age and gender match exactly, exactly one MBRSEL in right place	Age matches exactly, gender off, exactly one MBRSEL in right place	37
12	Age within one, gender matches exactly, exactly one MBRSEL in right place	Age matches exactly, gender off, exactly one MBRSEL in right place	3
13	Age within two, gender matches exactly, exactly one MBRSEL in right place	Age matches exactly, gender off, exactly one MBRSEL in right place	2
14	Age and gender match exactly, MBRSEL missing for all roster members	Age matches exactly, gender off, exactly one MBRSEL in right place	1
15	Age matches exactly, gender off, exactly one MBRSEL in right place	Age matches exactly, gender off, exactly one MBRSEL in right place	2
16	Age and gender match exactly, exactly one MBRSEL in right place	Age within one, gender matches exactly, MBRSEL missing for all roster members	38
17	Age within one, gender matches exactly, exactly one MBRSEL in right place	Age within one, gender matches exactly, MBRSEL missing for all roster members	9
18	Age within two, gender matches exactly, exactly one MBRSEL in right place	Age within one, gender matches exactly, MBRSEL missing for all roster members	1
19	Age and gender match exactly, MBRSEL missing for all roster members	Age within one, gender matches exactly, MBRSEL missing for all roster members	5
20	Age and gender match exactly, exactly one MBRSEL in right place	Age within 10, gender matches, exactly one MBRSEL in right place, excludes cases where MBRSEL could have been applied to one of closer age	140

Exhibit O.1 Rules for Determining Matching Pairs, in Priority Order (continued)

Priority	Rule		
	Pair Member A	**Pair Member B**	**Count**
21	Age within one, gender matches exactly, exactly one MBRSEL in right place	Age within 10, gender matches, exactly one MBRSEL in right place, excludes cases where MBRSEL could have been applied to one of closer age	15
22	Age within two, gender matches exactly, exactly one MBRSEL in right place	Age within 10, gender matches, exactly one MBRSEL in right place, excludes cases where MBRSEL could have been applied to one of closer age	4
23	Age and gender match exactly, MBRSEL missing for all roster members	Age within 10, gender matches, exactly one MBRSEL in right place, excludes cases where MBRSEL could have been applied to one of closer age	6
24	Age and gender match exactly, exactly one MBRSEL in right place	Age within 10, gender matches, MBRSEL missing for all roster members, excludes cases where one of closer age could have been selected	5
25	Age within 10, gender matches, MBRSEL missing for all roster members, excludes cases where one of closer age could have been selected, household size = 2	Age within 10, gender matches, MBRSEL missing for all roster members, excludes cases where one of closer age could have been selected, household size = 2	1
26	Age within one, gender matches exactly, exactly one MBRSEL in right place	Age within 10, gender matches, MBRSEL missing for all roster members, excludes cases where one of closer age could have been selected	5
27	Age and gender match exactly, exactly one MBRSEL in right place	Everything missing	19
28	Age within one, gender matches exactly, exactly one MBRSEL in right place	Everything missing	2
29	Age and gender match exactly, exactly one MBRSEL in right place	Gender and reported household sizes match exactly, age missing, MBRSEL missing for all roster members	2

Exhibit O.1 Rules for Determining Matching Pairs, in Priority Order (continued)

Priority	Rule		
	Pair Member A	**Pair Member B**	**Count**
30	Age and gender match exactly, MBRSEL missing for all roster members	Multiple matches on age and gender, MBRSEL missing for all roster members, only one match is from a domain of interest (parent-child)	1
31	Age and gender match exactly, exactly one MBRSEL in right place	Multiple matches on age and gender, MBRSEL missing for all roster members, only one match is from a domain of interest (spouse-spouse)	1
32	No match, but no relationship codes are missing, and none involve domains of interest	No match, but no relationship codes are missing, and none involve domains of interest	10
33	Age and gender match exactly, exactly one MBRSEL in right place	Age matches exactly, gender off, MBRSEL missing for all roster members	1
34	Age and gender match exactly, exactly one MBRSEL in right place	No match at all (often paired respondent is missing from roster)	35
35	Age within one, gender matches exactly, exactly one MBRSEL in right place	No match at all (often paired respondent is missing from roster)	8
36	Age within two, gender matches exactly, exactly one MBRSEL in right place	No match at all (often paired respondent is missing from roster)	0
37	Age and gender match exactly, MBRSEL missing for all roster members	No match at all (often paired respondent is missing from roster)	1
53	Age matches exactly, gender off, exactly one MBRSEL in right place	No match at all (often paired respondent is missing from roster)	1
54	No match at all	No match at all	8

O.2 Rules for identifying pair relationships among pairs

Table O2 summarizes the rules used to identify the pair relationships, using the relationship codes and questionnaire ages of the two pair members. Because the child (12 to 17)-parent and child (12 to 20)-parent relationships can be derived from relationships created using 12 to 14 year olds, 15 to 17 year olds, and 18 to 20 year olds, these latter relationships are the ones referenced in the rules. The variable PAIRREL, which is the last column of the table, identifies the pair relationship as defined by Table 6.1 in the main body of this report. As with the rules for identifying which members of the roster belong to the pair, these rules are given in priority order. In the headers, the moniker "A" refers to pair member A, and "B" refers to pair

member B. The relationship between A and B is described in the columns "A-B Relationship," from the perspective of pair member A ("B to A, according to A") and the perspective of pair member B ("A to B, according to B"). Any constraints on the pair members (other than FIPE3) are given in the columns "Constraint on A" and "Constraint on B." These constraints include age constraints, where a range of ages (e.g., 12-17) indicates that the value of the questionnaire edited age (AGE) is between the numbers given. Also in this column, "child" and "children" are defined as (a) roster member(s) with nonmissing ages smaller than 18. The question FIPE3 asks if the respondent is the parent of a selected 12 to 17 year old. The responses given in the table are either "yes" or "no." The column for RELMATCH indicates the quality of the match between pair members, as defined in Table 6.4 in the main body of this report. In the table, blank cells mean that no restrictions were placed on that variable to determine the pair relationship.

Exhibit O.2 Rules for Identifying Pair Relationships among Pairs

Priority	A-B Relationship		Constraint on A	Constraint on B	FIPE3 (A)	FIPE3 (B)	PAIR-REL	REL-MATCH
	B to A, according to A	A to B, according to B						
1	parent	child	12-14				1	1
	child	parent		12-14				
2	parent	child	15-17				2	1
	child	parent		15-17				
3	parent	child	18-20				3	1
	child	parent		18-20				
4	parent	child	21+				4	1
	child	parent		21+				
5	sibling	sibling	12-14	15-17			5	1
	sibling	sibling	15-17	12-14				
6	sibling	sibling	12-17	18-25			6	1
	sibling	sibling	18-25	12-17				
7	sibling	sibling	no constraints, after considering #5 & #6				7	1
8	spouse/partner	spouse/partner	>=1 child	>=1 child			8	1
9	spouse/partner	spouse/partner	0 children, no bad data	0 children, no bad data			9	1
10	spouse/partner	spouse/partner	>=1 child	0 children, some bad data			8	1.5
	spouse/partner	spouse/partner	0 children, some bad data	>=1 child				
11	spouse/partner	roommate/nonrelative	>=1 child both sides, equal number each side				8	3
	roommate/nonrelative	spouse/partner	>=1 child both sides, equal number each side					

Exhibit O.2 Rules for Identifying Pair Relationships among Pairs (continued)

Priority	A-B Relationship		Constraint on A	Constraint on B	FIPE3 (A)	FIPE3 (B)	PAIRREL	REL-MATCH
	B to A, according to A	A to B, according to B						
12	partner	partner	>=1 child	0 children, but other's children in household			8	12
	partner	partner	0 children, but other's children in household	>=1 child				
13	spouse/partner	spouse/partner	no constraints, after considering #8-#12				10	1
14	grandchild	grandparent					11	14
	grandparent	grandchild						
15	parent-in-law	child-in-law					12	15
	child-in-law	parent-in-law						
	other relative	other relative						
	roommate/boarder/ nonrelative	roommate/boarder/ nonrelative						
16	roommate/boarder/ other relative/ nonrelative/in-laws	roommate/boarder/ other relative/ nonrelative/in-laws					13	16
17	parent	missing	12-14				1	17
	missing	parent		12-14				
18	child	missing		12-14			1	18
	missing	child	12-14					
19	parent	missing	15-17				2	19
	missing	parent		15-17				
20	child	missing		15-17			2	20
	missing	child	15-17					
21	parent	missing	18-20				3	21
	missing	parent		18-20				
22	child	missing		18-20			3	22
	missing	child	18-20					

Exhibit O.2 Rules for Identifying Pair Relationships among Pairs (continued)

Priority	A-B Relationship		Constraint on A	Constraint on B	FIPE3 (A)	FIPE3 (B)	PAIRREL	REL-MATCH
	B to A, according to A	A to B, according to B						
23	parent	missing	21+				4	23
	missing	parent		21+				
24	child	missing		21+			4	24
	missing	child	21+					
25	sibling	missing	12-14	15-17			5	25
			15-17	12-14				
	missing	sibling	12-14	15-17				
			15-17	12-14				
26	sibling	missing	12-17	18-25			6	26
			18-25	12-17				
	missing	sibling	12-17	18-25				
			18-25	12-17				
27	sibling	missing	no constraints, after considering #24, #25				7	2
	missing	sibling	no constraints, after considering #24, #25					
28	spouse/partner	missing	>=1 child	no spouse in roster			8	28
	missing	spouse/partner	no spouse in roster	>=1 child				
29	spouse/partner	missing	0 children, no bad data	no spouse in roster			9	29
	missing	spouse/partner	no spouse in roster	0 children, no bad data				
30	spouse/partner	missing	after #27, #28, no constraints	no spouse in roster			10	30
	missing	spouse/partner	no spouse in roster	after #27, #28, no constraints				

Exhibit O.2 Rules for Identifying Pair Relationships among Pairs (continued)

Priority	A-B Relationship		Constraint on A	Constraint on B	FIPE3 (A)	FIPE3 (B)	PAIRREL	REL-MATCH
	B to A, according to A	A to B, according to B						
31	grandchild	missing	A at least 20 year older than B				11	2
	missing	grandparent						
	grandparent	missing	B at least 20 yrs older than A					
	missing	grandchild						
32	roommate/boarder/ other relative/ nonrelative/in-laws	missing			no		12	2
	missing	roommate/boarder/ other relative/ nonrelative/in-laws			no			
33	roommate/boarder/ other relative/ nonrelative/in-laws	missing					13	33
		roommate/boarder/ other relative/ nonrelative/in-laws						
34	child	nonmissing		12-14	yes		1	34
	nonmissing	parent		12-14	yes		1	
35	nonmissing	child	12-14			yes	1	35
	parent	nonmissing	12-14			yes	1	
36	child	nonmissing		15-17	yes		2	36
	nonmissing	parent		15-17	yes		2	
37	nonmissing	child	15-17			yes	2	37
	parent	nonmissing	15-17			yes	2	

Exhibit O.2 Rules for Identifying Pair Relationships among Pairs (continued)

Priority	A-B Relationship		Constraint on A	Constraint on B	FIPE3 (A)	FIPE3 (B)	PAIRREL	REL-MATCH
	B to A, according to A	A to B, according to B						
38	parent	roommate/boarder/ othr relative/ nonrelative	12-14			no	13	38
						missing	15	
	roommate/boarder/ other relative/ nonrelative	parent		12-14	no		13	
					missing		15	
39	parent	roommate/boarder/ other relative/ nonrelative	15-17			no	13	39
						missing	16	
	roommate/boarder/ other relative/ nonrelative	parent		15-17	no		13	
					missing		16	
40	parent	roommate/boarder/ other relative/ nonrelative	18-20				17	40
	roommate/boarder/ other relative/ nonrelative	parent		18-20			17	
41	parent	roommate/boarder/ other relative/ nonrelative	21+				18	41
	roommate/boarder/ other relative/ nonrelative	parent		21+			18	

Exhibit O.2 Rules for Identifying Pair Relationships among Pairs (continued)

Priority	A-B Relationship		Constraint on A	Constraint on B	FIPE3 (A)	FIPE3 (B)	PAIRREL	REL-MATCH
	B to A, according to A	A to B, according to B						
42	nonmissing not a sibling	child	12-14	21-75		no	13	42
			12-14, exactly one parent	21-75, exactly one spouse		missing	1	
			12-14, 0 or 2 parents, or B has 0 or 2 spouse	21-75, 0 or 2 spouses, or A has 0 or 2 parents		missing	15	
	child	nonmissing not a sibling	21-75	12-14	no		13	
			21-75, exactly one spouse	12-14, exactly one parent	missing		1	
			21-75, 0 or 2 spouses, or A has 0 or 2 parents	12-14, 0 or 2 parents, or B has 0 or 2 spouse	missing		15	
43	nonmissing not a sibling	child	15-17	24-75		no	13	43
			15-17, exactly one parent	24-75, exactly one spouse		missing	2	
			15-17, 0 or 2 parents, or B has 0 or 2 spouse	24-75, 0 or 2 spouses, or A has 0 or 2 parents		missing	16	
	child	nonmissing not a sibling	24-75	15-17	no		13	
			24-75, exactly one spouse	15-17, exactly one parent	missing		2	
			24-75, 0 or 2 spouses, or A has 0 or 2 parents	15-17, 0 or 2 parents, or B has 0 or 2 spouse	missing		16	

Exhibit O.2 Rules for Identifying Pair Relationships among Pairs (continued)

Priority	A-B Relationship		Constraint on A	Constraint on B	FIPE3 (A)	FIPE3 (B)	PAIRREL	REL-MATCH
	B to A, according to A	A to B, according to B						
44	nonmissing not a sibling	child	18-20, exactly one parent	27-75, exactly one spouse		missing	3	44
			18-20, 0 or 2 parents, or B has 0 or 2 spouse	27-75, 0 or 2 spouses, or A has 0 or 2 parents		missing	17	
	child	nonmissing not a sibling	27-75, exactly one spouse	18-20, exactly one parent	missing		3	
			27-75, 0 or 2 spouses, or A has 0 or 2 parents	18-20, 0 or 2 parents, or B has 0 or 2 spouse	missing		17	
45	nonmissing not a sibling	child	21+, exactly one parent	27-75, exactly one spouse		missing	4	45
			21+, 0 or 2 parents, or B has 0 or 2 spouse	27-75, 0 or 2 spouses, or A has 0 or 2 parents		missing	18	
	child	nonmissing not a sibling	27-75, exactly one spouse	21+, exactly one parent	missing		4	
			27-75, 0 or 2 spouses, or A has 0 or 2 parents	21+, 0 or 2 parents, or B has 0 or 2 spouse	missing		18	
46	spouse	sibling	one is 12-14, other, 15-17 both sides have parents				5	3
47	sibling	spouse	one is 12-14, other, 15-17 both sides have parents				5	3
48	spouse	sibling	one is 12-17, other, 18-25 both sides have parents				6	3
			one is 12-17, other, 18-25 both sides have spouses				6	3

Exhibit O.2 Rules for Identifying Pair Relationships among Pairs (continued)

Priority	A-B Relationship		Constraint on A	Constraint on B	FIPE3 (A)	FIPE3 (B)	PAIRREL	REL-MATCH
	B to A, according to A	A to B, according to B						
49	sibling	spouse	one is 12-17, other, 18-25 both sides have parents				6	3
			one is 12-17, other, 18-25 both sides have spouses				6	3
50	spouse	sibling	ages neither 12-14/15-17 nor 12-17/18-25; both sides have parents				7	3
			ages neither 12-14/15-17 nor 12-17/18-25; both sides have spouses				7	3
51	sibling	spouse	ages neither 12-14/15-17 nor 12-17/18-25; both sides have parents				7	3
			ages neither 12-14/15-17 nor 12-17/18-25; both sides have spouses				7	3
52	other relative	sibling	both sides have 2 parents; ages of oldest parents on either side differ by > 5 years; age of youngest parents on either side differ by > 5 years				13	3
53	sibling	other relative	both sides have 2 parents; ages of oldest parents on either side differ by > 5 years; age of youngest parents on either side differ by > 5 years				13	3
54	nonmissing, not child	sibling	15-17	12-14			19	54
55	sibling	nonmissing, not child	12-14	15-17			19	55
56	nonmissing, not parent	sibling	12-14	15-17			19	56
57	sibling	nonmissing, not parent	15-17	12-14			19	57
58	nonmissing, not child	sibling	18-25	12-17			20	58
59	sibling	nonmissing, not child	12-17	18-25			20	59
60	nonmissing, not parent	sibling	12-17	18-25			20	60
61	sibling	nonmissing, not parent	18-25	12-17			20	61

Exhibit O.2 Rules for Identifying Pair Relationships among Pairs (continued)

Priority	A-B Relationship		Constraint on A	Constraint on B	FIPE3 (A)	FIPE3 (B)	PAIRREL	REL-MATCH
	B to A, according to A	A to B, according to B						
Priority	**A-B Relationship**	**Constraint on A**	**Constraint on B**		**FIPE3 (A)**	**FIPE3 (B)**	**PAIRREL**	**REL-MATCH**
	B to A, according to A	**A to B, according to B**						
62	nonmissing, not child	sibling	ages neither 12-14/15-17 nor 12-17/18-25, A older than B				21	4
63	sibling	nonmissing, not child	ages neither 12-14/15-17 nor 12-17/18-25, B older than A				21	4
64	nonmissing, not parent	sibling	ages neither 12-14/15-17 nor 12-17/18-25, B older than A				21	4
65	sibling	nonmissing, not parent	ages neither 12-14/15-17 nor 12-17/18-25, A older than B				21	4
66	sibling	roommate, in-law, grandparent, grandchild, boarder, other relative, nonrelative	at least one is between 18 and 20			13	3	66
67	roommate, in-law, grandparent, grandchild, boarder, other relative, nonrelative	sibling	at least one is between 18 and 20			13	3	67
68	sibling	unusual in-law code	12-20	26 or over			13	68
69	unusual in-law code	sibling	26 or over	12-20			13	69
70	spouse/ partner	not a child, parent, or sibling	>=1 child aged < 18	no spouse			22	70
71	not a child, parent, or sibling	spouse/partner	no spouse	>=1 child aged < 18			22	71
72	spouse/partner	not a child, parent, or sibling	15 or over, 0 children, no bad data	15 or over, no spouse			23	72

Exhibit O.2 Rules for Identifying Pair Relationships among Pairs (continued)

Priority	A-B Relationship		Constraint on A	Constraint on B	FIPE3 (A)	FIPE3 (B)	PAIRREL	REL-MATCH
	B to A, according to A	A to B, according to B						
73	not a child, parent, or sibling	spouse/partner	15 or over, no spouse	15 or over, 0 children, no bad data			23	4
74	grandparent, grandchild	not grandparent, not grandchild					25	4
75	not grandparent, not grandchild	grandparent, grandchild					25	4
76	any codes	any codes	no constraints	no constraints			14	0

Appendix P

Conditions for Creating Household-Consistent Covariates

Appendix P

Conditions for Creating Household-Consistent Covariates

P.1 Household size

In Table P.1, blank entries indicate that no conditions were required for that set of variables. A variable followed by "A" in parentheses indicates that the variable corresponds to the value for pair member "A." A similar comment can be made with regard to the parenthetical "B." The reported household size variable is QD49, and the edited household size variable is TOTPEOP, which cannot differ from the raw variable by more than 1. The quality of roster counts are considered in the column "any roster missing?" The variables GOODAGEA and GOODAGEB are the total number of cases in the roster with valid ages. The variables that appear in the table are TGOODAGA and TGOODAGB, the total number of cases in the roster with valid ages, incorporating the minimum possible counts within the age categories 12 to 17, 18 to 25, 26 to 34, 35 to 49, and 50 and over. Finally, the variable used to describe the screener household size is SHHSIZE. The conditions used to create the variable HHSIZE resulted in no missing values for this variable, so that no imputation was required.

Exhibit P.1 Priority Rules Used to Create Household-Consistent Household Size

Priority	Relationship of QD49 (A) & QD49 (B)	Relationship of TOTPEOP (A) & TOTPEOP (B)	Relationships Involving TGOODAGA & TGOODAGB	Any Roster Missing?	Screener HHSIZE Characteristics	HHSIZE Equals:
1		equal, both > 1, both nonmissing				TOTPEOP (A)
2-7	equal, both > 1, both nonmissing	TOTPEOP (B) one more than TOTPEOP (A)	TGOODAGA<= QD49(A)	none in A		QD49 (A)
		TOTPEOP (A) one more than TOTPEOP (B)	TGOODAGB<= QD49(B)	none in B		QD49 (B)
		TOTPEOP (B) one more than TOTPEOP (A)	equal to each other, TGOODAGB<= TOTPEOP(B)		not equal to QD49(B)	TOTPEOP(B)
			TGOODAGB= TOTPEOP(B)		no condition	
		TOTPEOP (A) one more than TOTPEOP (B)	equal to each other, TGOODAGA<= TOTPEOP(A)		not equal to QD49(A)	TOTPEOP(A)
			TGOODAGA= TOTPEOP(A)		no condition	
6	equal, both > 1, both nonmissing	within one of each other			equal to QD49 (A) & (B)	SHHSIZE
7-10		A:missing or 1 B:not missing > 1			nonmissing, closer to QD49 (B) than TOTPEOP (B)	QD49 (B)
			TGOODAGB <= TOTPEOP (B)		nonmissing, TOTPEOP (B) is as close as QD49 (B)	TOTPEOP (B)
			TGOODAGB <= SHHSIZE		<== previous column	SHHSIZE
						TGOOD-AGB

Exhibit P.1 Priority Rules Used to Create Household-Consistent Household Size (continued)

Priority	Relationship of QD49 (A) & QD49 (B)	Relationship of TOTPEOP (A) & TOTPEOP (B)	Relationships Involving TGOODAGA & TGOODAGB	Any Roster Missing?	Screener HHSIZE Characteristics	HHSIZE Equals:
11-14		A:not missing, > 1 B:missing or 1			nonmissing, closer to QD49 (A) than TOTPEOP (A)	QD49 (A)
			TGOODAGB <= TOTPEOP (B)		nonmissing, TOTPEOP (A) is as close as QD49 (A)	TOTPEOP (A)
			TGOODAGB <= SHHSIZE		<== previous column	SHHSIZE
						TGOOD-AGB
15	both missing or 1	both missing or 1			nonmissing	SHHSIZE
16	not equal, both > 1, see next columns	TOTPEOP(A)=QD49 (A) or TOTPEOP(B) <= QD49 (A)			nonmissing equal to QD49 (A)	QD49(A)
17		TOTPEOP(A)=QD49 (B) or TOTPEOP(B) <= QD49 (B)			nonmissing equal to QD49 (B)	QD49(B)
18			TGOODAGA> GOODAGEA; TGOODAGB> GOODAGEB see SHHSIZE column	A: no B: no	nonmissing equal to TGOOD-AGA or TGOOD-AGB	SHHSIZE
19			TGOODAGA> GOODAGEA; TGOODAGB> GOODAGEB equal to each other	A: no B: no		TGOOD-AGA
20				A: no B: no	equal to the sum of maxima for each age group across pair members	SHHSIZE

Exhibit P.1 Priority Rules Used to Create Household-Consistent Household Size (continued)

Priority	Relationship of QD49 (A) & QD49 (B)	Relationship of TOTPEOP (A) & TOTPEOP (B)	Relationships Involving TGOODAGA & TGOODAGB	Any Roster Missing?	Screener HHSIZE Characteristics	HHSIZE Equals:
21-22				A: no B: no	>0, closer to QD49 (A) than QD49 (B)	QD49 (A)
					>0, closer to QD49 (B) than QD49 (A)	QD49 (B)
23-24	not equal, both > 1, see next columns			A: no B: no	equidistant between QD49 (A) and (B)	QD49 of oldest pair member
					equal to both QD49 (A) and QD49 (B)	SHHSIZE
25-28				fewer in A than in B	nonmissing, closer to QD49 (A) than QD49 (B)	QD49 (A)
					nonmissing, closer to QD49 (B) than QD49 (A)	QD49 (B)
					equidistant between QD49 (A) and (B)	QD49 corresponding to oldest pair member
					equal to both QD49 (A) and QD49 (B)	SHHSIZE
29-30				fewer in B than in A	nonmissing, closer to QD49 (A) than QD49 (B)	QD49 (A)
					nonmissing, closer to QD49 (B) than QD49 (A)	QD49 (B)
31-32	not equal, both > 1, see next columns			A: no B: no	equidistant between QD49 (A) and (B)	QD49 corresponding to oldest pair member
					equal to both QD49 (A) and QD49 (B)	SHHSIZE

P.2. Age variables

Table P.2 illustrates the hierarchical conditions ("priorities") used to create a new household-consistent 12 to 17 age group count; similar conditions are used for the 18 to 25, 26 to 34, 35 to 49, and 50+ age groups. In this table, blank entries indicate that no conditions were required for that set of variables. As with the previous set of tables, a variable followed by "A" (either in parentheses or not) indicates that the variable corresponds to the value for pair member "A." A similar comment can be made with regard to the "B." As stated earlier, the variables GOODAGEA and GOODAGEB are the total number of cases in the roster with valid ages, and the variables TGOODAGA and TGOODAGB are also the total number of cases in the roster with valid ages, but if the original adjusted count is less than the minimum required, the original count is replaced by the minimum within the age categories 12 to 17, 18 to 25, 26 to 34, 35 to 49, and 50 and over. As noted in Section 6.2, these counts are adjusted so that the roster ages match what was entered in each pair member's questionnaire. Hence, AGE1217A is the adjusted count of 12 to 17 year olds for pair member A, and AGE1217B is the adjusted count of 12 to 17 year olds for pair member B. If AGE1217A or AGE1217B is less than the minimum possible, the count is replaced by the minimum, which is given by TAG1217A and TAG1217B respectively. Otherwise, AGE1217A and TAG1217A are equivalent, as are AGE1217B and TAG1217B. The sum of AGE011A, AGE1217A, AGE1825A, AGE2634A, AGE3549A, and AGE50PA is GOODAGEA. Similarly, the sum of AGE011A, TAG1217A, TAG1825A, TAG2634A, TAG3549A, and TAG50PA is TGOODAGA. The same can be said for GOODAGEB and TGOODAGB. The final 12 to 17 age count is denoted by AGE1217. The screener age count, denoted by SAGE1217, is only used if the age counts in each pair member's roster cannot conform to the minimum necessary, or are otherwise not possible to incorporate. If, after all edits, the count for AGE1217 is missing but the count for other age groups are not, and the counts for the 0 to 11 age group are the same for both pair members, then the sum of the counts for the other age groups, plus the minimum possible for AGE1217, are given by EXC1217. As a final check, if the age group counts do not equal HHSIZE, and the count for the pair members are unequal, then the count is set to missing. This occurs for 10 cases in 1999.

Exhibit P.2 Priority Rules Used to Create Household-Consistent Age Variables

Priority	Relationships Involving TOTPEOP, GOODAGE, and HHSIZE	Relationships Involving AGE1217A, AGE1217B	Relationships Involving Other Age Groups	Relationships Involving Screener Counts	Quality of Roster Measures	AGE1217 Equals:
1-9	GOODAGEA equals GOODAGEB, GOODAGEA equals TOTPEOPA, GOODAGEB equals TOTPEOPB GOODAGEB equals HHSIZE all nonmissing, all > 1	AGE1217A < minimum, AGE1217B >= minimum				AGE1217B
		AGE1217B < minimum AGE1217A >= minimum				AGE1217A
		AGE1217A < minimum AGE1217B < minimum		screener count >= minimum		SAGE1217
		AGE1217A equals AGE1217B; both >= minimum	Another count except 12-17 < minimum			AGE1217A
		AGE1217A not equal to AGE1217B; both >= minimum	AGE1825A < minimum; AGE1825B >= minimum			AGE1217 B
			AGE1825B < minimum; AGE1825A >= minimum			AGE1217A
			Another count except 12-17 < minimum		fewer roster entries missing in A than B	AGE1217A
					fewer roster entries missing in B than A	AGE1217 B
					same # missing roster entries both sides; at least one of match measures 1, 2, and/or 8 (A)	AGE1217A

Exhibit P.2 Priority Rules Used to Create Household-Consistent Age Variables (continued)

Priority	Relationships Involving TOTPEOP, GOODAGE, and HHSIZE	Relationships Involving AGE1217A, AGE1217B	Relationships Involving Other Age Groups	Relationships Involving Screener Counts	Quality of Roster Measures	AGE1217 Equals:
10-19	GOODAGEA equals GOODAGEB, GOODAGEA equals TOTPEOPA, GOODAGEB equals TOTPEOPB GOODAGEB equals HHSIZE all nonmissing, all > 1	AGE1217A not equal to AGE1217B; both >= minimum	Another count except 12-17 < minimum		same # missing roster entries both sides; at least one of match measures 1, 2, and/or 8 (B)	AGE1217B
					same # missing roster entries both sides; Age (A) >= Age (B)	AGE1217A
					same # missing roster entries both sides; Age (B) > Age (A)	AGE1217B
20		AGE1217A equals AGE1217B	All other counts equal across pair members			AGE1217A
21-37		At least one age group has an unequal count between pair members			no missing roster entries on either side; A: match measure 1; B: not 1	AGE1217A
					no missing roster entries on either side; B: match measure 1; A: not 1	AGE1217B
					no missing roster entries on either side; same pattern of match measures either side	

Exhibit P.2 Priority Rules Used to Create Household-Consistent Age Variables (continued)

Priority	Relationships Involving TOTPEOP, GOODAGE, and HHSIZE	Relationships Involving AGE1217A, AGE1217B	Relationships Involving Other Age Groups	Relationships Involving Screener Counts	Quality of Roster Measures	AGE1217 Equals:
					no missing roster entries on either side; A older than B	AGE1217A
38-42	GOODAGEA equals GOODAGEB, GOODAGEA equals TOTPEOPA, GOODAGEB equals TOTPEOPB GOODAGEB equals HHSIZE all nonmissing, all > 1	At least one age group has an unequal count between pair members			no missing roster entries on either side; B older than A	AGE1217B
					fewer roster entries missing on A side than B	AGE1217A
					fewer roster entries missing on B side than A	AGE1217B
					both sides missing equal number of roster entries; B as old or older than A	AGE1217B
					both sides missing equal number of roster entries; A older than B	AGE1217A

Exhibit P.2 Priority Rules Used to Create Household-Consistent Age Variables (continued)

Priority	Relationships Involving TOTPEOP, GOODAGE, and HHSIZE	Relationships Involving AGE1217A, AGE1217B	Relationships Involving Other Age Groups	Relationships Involving Screener Counts	Quality of Roster Measures	AGE1217 Equals:
43-46	GOODAGEA equals TOTPEOPA, GOODAGEB equals TOTPEOPB GOODAGEA equals HHSIZE, GOODAGEB not equal to HHSIZE	AGE1217A < minimum, AGE1217B = minimum				AGE1217B
		AGE1217B < minimum AGE1217A = minimum				AGE1217A
		AGE1217A < minimum AGE1217B < minimum		screener count >= minimum		SAGE-1217
		AGE1217A equals AGE1217B; both >= minimum	Another A count except 12-17 < minimum			AGE1217A
47-50	GOODAGEA equals TOTPEOPA, GOODAGEB equals TOTPEOPB GOODAGEA equals HHSIZE, GOODAGEB not equal to HHSIZE	AGE1217A not equal to AGE1217B; both >= minimum	AGE1825A < minimum; AGE1825B >= minimum			AGE1217 B
			AGE1825B < minimum; AGE1825A >= minimum			AGE1217A
			Another A count except 12-17 < minimum		fewer roster entries missing in A than B	AGE1217A
					fewer roster entries missing in B than A	AGE1217 B
52-56	GOODAGEA equals TOTPEOPA, GOODAGEB equals TOTPEOPB GOODAGEB equals HHSIZE, GOODAGEA not equal to HHSIZE	AGE1217A < minimum, AGE1217B = minimum				AGE1217B
		AGE1217B < minimum AGE1217A = minimum				AGE1217A
		AGE1217A < minimum AGE1217B < minimum		screener count >= minimum		SAGE-1217
		AGE1217A equals AGE1217B; both >= minimum	Another B count except 12-17 < minimum			AGE1217B

Exhibit P.2 Priority Rules Used to Create Household-Consistent Age Variables (continued)

Priority	Relationships Involving TOTPEOP, GOODAGE, and HHSIZE	Relationships Involving AGE1217A, AGE1217B	Relationships Involving Other Age Groups	Relationships Involving Screener Counts	Quality of Roster Measures	AGE1217 Equals:
		AGE1217A not equal to AGE1217B; both >= minimum	AGE1825A < minimum; AGE1825B >= minimum			AGE1217B
57-59	GOODAGEA equals TOTPEOPA, GOODAGEB equals TOTPEOPB GOODAGEB equals HHSIZE, GOODAGEA not equal to HHSIZE	AGE1217A not equal to AGE1217B; both >= minimum	AGE1825B < minimum; AGE1825A >= minimum			AGE1217A
			Another B count except 12-17 < minimum		fewer roster entries missing in A than B	AGE1217A
					fewer roster entries missing in B than A	AGE1217 B
60	TGOODAGA equals HHSIZE					TAG1217A
61	TGOODAGB equals HHSIZE					TAG1217B
62	SHHSIZE equals HHSIZE	AGE1217A, AGE1217B <= SAGE1217		AGE1217A & B <= SAGE1217		SAGE-1217
	SHHSIZE equals HHSIZE, HHSIZE equals EXC1217	AGE1217 missing	other counts not missing, AGE011A equals AGE011B			MIN1217
63	Previous conditions for HHSIZE, TOTPEOP, GOODAGE, not met			AGE1217A equals SAGE1217		AGE1217A
				AGE1217B equals SAGE1217		AGE1217B

Appendix Q

Multiplicity and Household Count Model Summaries

Appendix Q

Multiplicity and Household Count Model Summaries

Q.1 Introduction

The exhibits in this appendix list the covariates used in all the models that were run to impute missing values in the pair relationship, multiplicity, and household count variables. For each variable or set of variables to which the predictive mean neighborhood (PMN) imputation method was applied, three models were run: one to adjust the weights for item nonresponse (response propensity models), and a second and third to calculate predictive means. In the second model, household composition was represented by the household size variable, HHSIZE, and in the third, household composition was represented by the household composition age count variables. Imputation was sometimes performed within separate model groups, so that separate exhibits are required for those model groups.

Section Q.2 deals with the pair relationship variables; Section Q.3 deals with the multiplicity variables; and Section Q.4 deals with the household-level person count variables. In the exhibits, when an asterisk "*" is given, it represents an interaction between two variables and not multiplication. In addition, when the initialism "MSA" is used, it represents "metropolitan statistical area." Finally, these models were at a pair level, whereas some of the variables in the models were at a person level. To differentiate which respondent the person-level variable applied to, the variable label is followed by a parenthetical "older" or "younger" to refer to the variable corresponding to the older or younger respondent, respectively. If the respondents in the pair were the same age, one of the respondents was randomly selected to be "older" or "younger."

Q.2 Pair Relationship Variables

Exhibit Q.1 Model Summaries (Pair Relationships)

Model Group	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model	
		Including Household Size	Not Including Household Size
0 (12-14, 12-14)	Race (older), Sex (older), Sex (younger), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment, Household Size (HHSIZE)	Race (older), Sex (older), Sex (younger), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment, Household Size (HHSIZE)	Number in Household Aged 0-11, Number in Household Aged 12-17, Number in Household Aged 18-25, Number in Household Aged 26-34, Number in Household Aged 35-49, Number in Household Aged 50+, Race (older), Sex (older), Sex (younger), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment
1 (12-14, 15-17)	Race (older), Sex (older), Sex (younger), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Owner-Occupied Households in Segment, Household Size (HHSIZE)	Race (older), Sex (older), Sex (younger), Education (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment, Household Size (HHSIZE)	Number in Household Aged 0-11, Number in Household Aged 12-17, Number in Household Aged 18-25, Number in Household Aged 26-34, Number in Household Aged 35-49, Number in Household Aged 50+, Age Category (older), Race (older), Sex (older), Sex (younger), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment
2 (12,14, 18-25)	Age Category (older), Race (older), Sex (older), Sex (younger), Education (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Owner-Occupied Households in Segment, Household Size (HHSIZE)	Age Category (older), Race (older), Sex (older), Sex (younger), Education (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment, Household Size (HHSIZE)	Number in Household Aged 0-11, Number in Household Aged 12-17, Number in Household Aged 18-25, Number in Household Aged 26-34, Number in Household Aged 35-49, Number in Household Aged 50+, Age Category (older), Race (older), Sex (older), Sex (younger), Education (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment

Exhibit Q.1 Model Summaries (Pair Relationships) (continued)

Model Group	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model	
		Including Household Size	**Not Including Household Size**
3 (15-17, 15-17)	Race (older), Sex (older), Sex (younger), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment, Household Size (HHSIZE)	Race (older), Sex (older), Sex (younger), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment, Household Size (HHSIZE)	Number in Household Aged 0-11, Number in Household Aged 12-17, Number in Household Aged 18-25, Number in Household Aged 26-34, Number in Household Aged 35-49, Number in Household Aged 50+, Race (older), Sex (older), Sex (younger), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment
4 (15-17, 18-25)	Age Category (older), Race (older), Sex (older), Sex (younger), Education (older), Marital Status (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment, Household Size (HHSIZE)	Age Category (older), Race (older), Sex (older), Sex (younger), Marital Status (older), Education (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment, Household Size (HHSIZE)	Number in Household Aged 0-11, Number in Household Aged 12-17, Number in Household Aged 18-25, Number in Household Aged 26-34, Number in Household Aged 35-49, Number in Household Aged 50+, Age Category (older), Race (older), Sex (older), Sex (younger), Marital Status (older), Education (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment
5 (18-20, 18-25)	Age Category (older), Race (older), Sex (older), Sex (younger), Marital Status (older), Marital Status (younger), Education (older), Education (younger), Employment (older),Employment (younger), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment, Household Size (HHSIZE)	Age Category (older), Race (older), Sex (older), Marital Status (older),Marital Status (younger), Education (older), Employment (older), Employment (younger), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment	Number in Household Aged 12-17, Age Category (older), Race (older), Sex (older), Sex (younger), Marital Status (older), Education (older), Employment (older), Employment (younger), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment

Exhibit Q.1 Model Summaries (Pair Relationships) (continued)

Model Group	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model	
		Including Household Size	Not Including Household Size
6 (21-25, 21-25)	Race (older), Sex (older), Sex (younger), Employment (older), Employment (younger), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment, Household Size (HHSIZE)	Race (older), Sex (older), Employment (older), Employment (younger), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment, Household Size (HHSIZE)	Number in Household Aged 26-34, Race (older), Sex (older), Sex (younger), Employment (older), Employment (younger), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment
7 (12-14, 26+)	Age Category (older), Race (older), Sex (older), Sex (younger), Marital Status (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Owner-Occupied Households in Segment, Household Size (HHSIZE)	Age Category (older), Race (older), Sex (older), Sex (younger), Marital Status (older), Education (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment, Household Size (HHSIZE)	Number in Household Aged 0-11, Number in Household Aged 12-17, Number in Household Aged 18-25, Number in Household Aged 26-34, Number in Household Aged 35-49, Number in Household Aged 50+, Age Category (older), Race (older), Sex (older), Sex (younger), Marital Status (older), Education (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment
8 (15-17, 26+	Age Category (older), Race (older), Sex (older), Sex (younger), Education (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Household Size (HHSIZE)	Age Category (older), Race (older), Sex (older), Sex (younger), Marital Status (older), Education (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment, Household Size (HHSIZE)	Number in Household Aged 0-11, Number in Household Aged 12-17, Number in Household Aged 18-25, Number in Household Aged 26-34, Number in Household Aged 35-49, Number in Household Aged 50+, Age Category (older), Race (older), Sex (older), Sex (younger), Marital Status (older), Education (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment

Exhibit Q.1 Model Summaries (Pair Relationships) (continued)

Model Group	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model	
		Including Household Size	**Not Including Household Size**
9 (18-20, 26+)	Age Category (older), Race (older), Sex (older), Sex (younger), Marital Status (older), Education (older), Education (younger), Employment (older), Employment (younger), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment, Household Size (HHSIZE)	Age Category (older), Race (older), Sex (older), Sex (younger), Marital Status (older), Education (older), Employment (older), Employment (younger), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment, Household Size (HHSIZE)	Number in Household Aged 0-11, Number in Household Aged 12-17, Number in Household Aged 18-25, Age Category (older), Race (older), Sex (older), Sex (younger), Education (older), Employment (older), Employment (younger), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment
10 (21+, 26+)	Age Category (older), Race (older), Sex (older), Sex (younger), Marital Status (older), Employment (older), Employment (younger), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment, Household Size (HHSIZE)	Race (older), Sex (older), Employment (older), Employment (younger), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment, Household Size (HHSIZE)	Race (older), Sex (older), Employment (older), Employment (younger), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment

Q.3 Multiplicities

Exhibit Q.2 Model Summaries (Multiplicities)

Pair Domain	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model	
		Including Household Size	Not Including Household Size
Parent-child (12-20) parent focus	Race (older), Sex (older), Sex (younger), Marital Status (older), Education (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment, Household Size (HHSIZE)	Age Category (older), Race (older), Sex (older), Sex (younger), Marital Status (older), Education (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment, Household Size (HHSIZE)	Number in Household Aged 0-11, Number in Household Aged 12-17, Number in Household Aged 18-25, Number in Household Aged 26-34, Number in Household Aged 35-49, Number in Household Aged 50+, Age Category (older), Race (older), Sex (older), Sex (younger), Marital Status (older), Education (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment
Parent-child (12-20) child focus	Race (older), Sex (older), Sex (younger), Marital Status (older), Education (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment, Household Size (HHSIZE)	Age Category (older), Race (older), Sex (older), Sex (younger), Marital Status (older), Education (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment, Household Size (HHSIZE)	Number in Household Aged 0-11, Number in Household Aged 12-17, Number in Household Aged 18-25, Number in Household Aged 26-34, Number in Household Aged 35-49, Number in Household Aged 50+, Age Category (older), Race (older), Sex (older), Sex (younger), Marital Status (older), Education (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment
Sibling (12-14) Sibling (15-17) Older Sibling Focus	Race (older), Sex (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment, Household Size (HHSIZE)	Race (older), Sex (older), Sex (younger), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment, Household Size (HHSIZE)	Number in Household Aged 0-11, Number in Household Aged 12-17, Number in Household Aged 18-25, Number in Household Aged 26-34, Number in Household Aged 35-49, Number in Household Aged 50+, Race (older), Sex (older), Sex (younger), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment

Exhibit Q.2 Model Summaries (Multiplicities) (continued)

Pair Domain	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model	
		Including Household Size	Not Including Household Size
Sibling (12-14) Sibling (15-17) Younger Sibling Focus	Race (older), Sex (older), Sex (younger), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment, Household Size (HHSIZE)	Race (older), Sex (older), Sex (younger), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment, Household Size (HHSIZE)	Number in Household Aged 0-11, Number in Household Aged 12-17, Number in Household Aged 18-25, Number in Household Aged 26-34, Number in Household Aged 35-49, Number in Household Aged 50+, Race (older), Sex (older), Sex (younger), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment
Sibling (12-17) Sibling (18-25) Older Sibling Focus	Race (older), Sex (older), Education (older), Marital Status (older), Employment (older), MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment, Household Size (HHSIZE)	Age Category (older), Race (older), Sex (older), Sex (younger), Marital Status (older), Education (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment, Household Size (HHSIZE)	Number in Household Aged 0-11, Number in Household Aged 12-17, Number in Household Aged 18-25, Number in Household Aged 26-34, Number in Household Aged 35-49, Number in Household Aged 50+, Age Category (older), Race (older), Sex (older), Sex (younger), Marital Status (older), Education (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment
Sibling (12-17) Sibling (18-25) Younger Sibling Focus	Age Category (older), Race (older), Sex (older), Sex (younger), Marital Status (older), Education (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment	Age Category (older), Race (older), Sex (older), Marital Status (older), Education (older), Employment (older), Employment (younger), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment, Household Size (HHSIZE)	Number in Household Aged 0-11, Number in Household Aged 12-17, Number in Household Aged 18-25, Number in Household Aged 26-34, Number in Household Aged 35-49, Number in Household Aged 50+, Age Category (older), Race (older), Sex (older), Sex (younger), Marital Status (older), Education (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment

Q.4 Household-Level Person Counts

Exhibit Q.3 Model Summaries (Household-Level Person Counts of Pair Domains when Respondent is in a Responding Pair)

Model Group	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model	
		Including Household Size	Not Including Household Size
Parent-child (12-20) child focus, both pair members < 18	Race (older), Sex (older), Sex (younger), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment, Household Size (HHSIZE)	Race (older), Sex (older), Sex (younger), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment, Household Size (HHSIZE)	Number in Household Aged 0-11, Number in Household Aged 12-17, Number in Household Aged 18-25, Number in Household Aged 26-34, Number in Household Aged 35-49, Number in Household Aged 50+, Race (older), Sex (older), Sex (younger), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment
Parent-child (12-20) child focus, at least one pair member older than 18	Race (older), Sex (older), Sex (younger), Marital Status (older), Education (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Owner-Occupied Households in Segment, Household Size (HHSIZE)	Race (older), Sex (older), Sex (younger), Marital Status (older), Education (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment, Household Size (HHSIZE)	Number in Household Aged 0-11, Number in Household Aged 12-17, Number in Household Aged 18-25, Number in Household Aged 26-34, Number in Household Aged 35-49, Number in Household Aged 50+, Race (older), Sex (older), Sex (younger), Marital Status (older), Education (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment
Parent-child (12-20) parent focus, both pair members <18	Race (older), Sex (older), Sex (younger), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Owner-Occupied Households in Segment, Household Size (HHSIZE)	Race (older), Sex (older), Sex (younger), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment, Household Size (HHSIZE)	Number in Household Aged 0-11, Race (older), Sex (older), Sex (younger), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment

Exhibit Q.3 Model Summaries (Household-Level Person Counts of Pair Domains when Respondent is in a Responding Pair) (continued)

Model Group	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model	
		Including Household Size	Not Including Household Size
Parent-child (12-20) parent focus, at least one pair member older than 18	Race (older), Sex (older), Sex (younger), Education (older), Marital Status (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment, Household Size (HHSIZE)	Race (older), Sex (older), Sex (younger), Education (older), Marital Status (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment, Household Size (HHSIZE)	Number in Household Aged 0-11, Race (older), Sex (older), Sex (younger), Education (older), Marital Status (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment
Sibling (12-14) Sibling (15-17), Older Sibling Focus, both pair members <18	Race (older), Sex (older), Sex (younger), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment, Household Size (HHSIZE)	Age Category (older), Race (older), Sex (older), Sex (younger), Marital Status (older), Education (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment, Household Size (HHSIZE)	Number in Household Aged 0-11, Number in Household Aged 12-17, Number in Household Aged 18-25, Number in Household Aged 26-34, Number in Household Aged 35-49, Number in Household Aged 50+, Race (older), Sex (older), Sex (younger), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment
Sibling (12-14) Sibling (15-17), Older Sibling Focus, at least one pair member older than 18	Race (older), Sex (older), Sex (younger), Marital Status (older), Education (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment, Household Size (HHSIZE)	Race (older), Sex (older), Marital Status (older), Education (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment	Number in Household Aged 0-11, Number in Household Aged 12-17, Number in Household Aged 18-25, Number in Household Aged 26-34, Number in Household Aged 35-49, Number in Household Aged 50+, Race (older), Sex (older), Sex (younger), Marital Status (older), Education (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment

Exhibit Q.3 Model Summaries (Household-Level Person Counts of Pair Domains when Respondent is in a Responding Pair) (continued)

Model Group	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model	
		Including Household Size	Not Including Household Size
Sibling (12-17) Sibling (18-25), Older Sibling Focus, both pair members <18	Race (older), Sex (older), Sex (younger), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Household Size (HHSIZE)	Race (older), Sex (older), Employment (older), Employment (younger), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment, Household Size (HHSIZE)	Number in Household Aged 0-11, Number in Household Aged 12-17, Number in Household Aged 18-25, Number in Household Aged 26-34, Number in Household Aged 35-49, Number in Household Aged 50+, Race (older), Sex (older), Sex (younger), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment
Sibling (12-17) Sibling (18-25), Older Sibling Focus, at least one pair member older than 18	Race (older), Sex (older), Sex (younger), Education (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Owner-Occupied Households in Segment, Household Size (HHSIZE)	Age Category (older), Race (older), Sex (older), Sex (younger), Marital Status (older), Education (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment, Household Size (HHSIZE)	Number in Household Aged 0-11, Number in Household Aged 12-17, Number in Household Aged 18-25, Number in Household Aged 26-34, Number in Household Aged 35-49, Number in Household Aged 50+, Race (older), Sex (older), Sex (younger), Marital Status (older), Education (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment
Spouse-spouse, both pair members < 18	Race (older), Sex (older), Sex (younger), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Household Size (HHSIZE)	Age Category (older), Race (older), Sex (older), Sex (younger), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment	Number in Household Aged 0-11, Race (older), Sex (older), Sex (younger), Marital Status (older), Education (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment

Exhibit Q.3 Model Summaries (Household-Level Person Counts of Pair Domains when Respondent is in a Responding Pair) (continued)

Model Group	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model	
		Including Household Size	Not Including Household Size
Spouse-spouse, at least one pair member older than 18	Age Category (older), Race (older), Sex (older), Sex (younger), Marital Status (older), Education (older), Employment (older), Employment (younger), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment, Household Size (HHSIZE)	Age Category (older), Race (older), Sex (older), Sex (younger), Marital Status (older), Education (older), Employment (older), Employment (younger), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment, Household Size (HHSIZE)	Number in Household Aged 0-11, Race (older), Sex (older), Sex (younger), Marital Status (older), Education (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment
Spouse-spouse with children, both pair members <18	Race (older), Sex (older), Sex (younger), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment, Household Size (HHSIZE), Number of Spouse-Spouse	Race (older), Sex (older), Sex (younger), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment, Household Size (HHSIZE)	Number in Household Aged 0-11, Number in Household Aged 12-17, Number in Household Aged 18-25, Number in Household Aged 26-34, Number in Household Aged 35-49, Number in Household Aged 50+, Race (older), Sex (older), Sex (younger), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment
Spouse-spouse with children, at least one pair member older than 18	Race (older), Sex (older), Sex (younger), Marital Status (older), Education (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment, Household Size (HHSIZE)	Race (older), Sex (older), Sex (younger), Marital Status (older), Education (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Owner-Occupied Households in Segment	Race (older), Sex (older), Sex (younger), Marital Status (older), Education (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Owner-Occupied Households in Segment

Exhibit Q.4 Model Summaries (Household-Level Person Counts of Pair Domains when Respondent is not in a Responding Pair)

Model Group	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model	
		Including Household Size	Not Including Household Size
Parent-child (12-20) child focus, < 18	Race, Sex, Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment, Household Size (HHSIZE)	Race, Sex, Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment, Household Size (HHSIZE)	Number in Household Aged 0-11, Number in Household Aged 12-17, Number in Household Aged 18-25, Number in Household Aged 26-34, Number in Household Aged 35-49, Number in Household Aged 50+, Race, Sex, Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment
Parent-child (12-20) child focus, older than 18	Race, Sex, Marital Status, Education, Employment, Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Owner-Occupied Households in Segment, Household Size (HHSIZE)	Age Category, Race, Sex, Marital Status, Education, Employment, Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment, Household Size (HHSIZE)	Number in Household Aged 0-11, Number in Household Aged 12-17, Number in Household Aged 18-25, Number in Household Aged 26-34, Number in Household Aged 35-49, Number in Household Aged 50+, Race, Sex, Marital Status, Education, Employment, Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment
Parent-child (12-20) parent focus, <18	Race, Sex, Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Owner-Occupied Households in Segment, Household Size (HHSIZE)	Race, Sex, Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment	Race, Sex, Census Region, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment
Parent-child (12-20) parent focus, older than 18	Race, Sex, Education, Marital Status, Employment, Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment, Household Size (HHSIZE)	Race, Sex, Education, Employment, Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment	Number in Household Aged 0-11, Race, Sex, Employment, Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment

Exhibit Q.4 Model Summaries (Household-Level Person Counts of Pair Domains when Respondent is not in a Responding Pair) (continued)

Model Group	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model	
Sibling (12-14) Sibling (15-17), Older Sibling Focus, <18	Race, Sex, Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment, Household Size (HHSIZE)	Race, Sex, Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment, Household Size (HHSIZE)	Number in Household Aged 0-11, Number in Household Aged 12-17, Number in Household Aged 18-25, Number in Household Aged 26-34, Number in Household Aged 35-49, Number in Household Aged 50+, Race, Sex, Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment
Sibling (12-14) Sibling (15-17), Older Sibling Focus, older than 18	Race, Sex, Marital Status, Education, Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment, Household Size (HHSIZE)	Age Category, Race, Sex, Marital Status, Education, Employment, Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment	Number in Household Aged 0-11, Number in Household Aged 12-17, Number in Household Aged 18-25, Number in Household Aged 26-34, Number in Household Aged 35-49, Number in Household Aged 50+, Race, Sex, Marital Status, Education, Employment, Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment
Sibling (12-17) Sibling (18-25), Older Sibling Focus, <18	Race, Sex, Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Household Size (HHSIZE)	Race, Sex, Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment, Household Size (HHSIZE)	Number in Household Aged 0-11, Number in Household Aged 12-17, Number in Household Aged 18-25, Number in Household Aged 26-34, Number in Household Aged 35-49, Number in Household Aged 50+, Race, Sex, Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment

Exhibit Q.4 Model Summaries (Household-Level Person Counts of Pair Domains when Respondent is not in a Responding Pair) (continued)

Model Group	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model	
Sibling (12-17) Sibling (18-25), Older Sibling Focus, older than 18	Race, Sex, Marital Status, Education, Employment, Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Owner-Occupied Households in Segment, Household Size (HHSIZE)	Age Category, Race, Sex, Marital Status, Education, Employment, Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment, Household Size (HHSIZE)	Number in Household Aged 0-11, Number in Household Aged 12-17, Number in Household Aged 18-25, Number in Household Aged 26-34, Number in Household Aged 35-49, Number in Household Aged 50+, Race, Sex, Marital Status, Education, Employment, Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment
Spouse-spouse, < 18	Race, Sex, Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment, Household Size (HHSIZE)	Sex, Census Region	Sex, Census Region
Spouse-spouse, older than 18	Race, Sex, Marital Status, Education, Employment, Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment, Household Size (HHSIZE)	Age Category, Race, Sex, Education, Employment, Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment	Race, Sex, Education, Employment, Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment
Spouse-spouse with children	Race, Sex, Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment, Household Size (HHSIZE)	Race, Sex, Marital Status, Education, Employment, Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment, Household Size (HHSIZE)	Number in Household Aged 0-11, Number in Household Aged 12-17, Number in Household Aged 18-25, Number in Household Aged 26-34, Number in Household Aged 35-49, Number in Household Aged 50+, Race, Sex, Marital Status, Education, Employment, Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment

Appendix R

Conditions Used for Reconciling Differing Multiplicity Counts between Pair Members

Appendix R

Conditions Used for Reconciling Differing Multiplicity Counts between Pair Members

R.1 Introduction

In order to determine multiplicity counts, counts were obtained from each pair member. The count from the pair member who was the focus member of the domain is considered the direct count, and the count from the other pair member is considered the indirect count. Typically, these counts were in agreement, and the determination of the final multiplicity count was straightforward, provided both rosters did not have bad data codes. The strategy was also usually clear if one pair member had bad data in the household roster; the count from the pair member with good data was usually preferred in those cases. If the bad data was limited to bad relationship codes, then the member with good data was only selected if substituting the appropriate relationship codes for the bad data codes would have given a total that was equal to the count from the pair member with good data. There were instances where bad data codes existed in the roster, and this condition did not apply. There were other exceptions as well. Finally, there were instances where neither pair member had bad data in their rosters, yet their counts still disagreed. In this appendix, the rules that were used to reconcile these disagreeing counts are outlined.

R.2 Parent-child counts

For parent-child counts, the screener and the FIPE3 variable were used to help reconcile disagreeing counts. The rules follow below, separated by the member of focus:

Parent-child pairs, child focus. The multiplicity counts in this domain reflected the selected child's parents, and were limited to have values of 1 or 2. If neither side had bad relationship codes, and the direct count was 2 while the indirect count was 1, the following rules applied:

1. The direct count might have exceeded the indirect count because one parent had left or entered the household between interviews. In this case, the ages in the rosters were matched to the screener roster to determine which count to believe.

2. The direct count might have exceeded the indirect count because the selected parent did not consider the other "parent" a spouse or live-in partner. If the pair relationship was not imputed, the indirect count was selected. However, if the pair relationship was imputed and the older pair member called the younger pair member a child, then the older pair member considered the child's "true" parent as not a spouse or live-in partner, even though he/she claimed the "true" parent's children. In this case, the direct count was used (the child's adjusted count).

If the direct count was 1 but the indirect count was 2, the child only listed one parent, but the parent lists a spouse (a "stepparent") or live-in partner in the household roster. The following rules applied:

1. The indirect count might exceed the direct count because the selected child did not accept a stepparent or live-in partner as his/her parent. If this stepparent or live-in partner was the other respondent selected, we determined this was a child-parent pair based on the response of the "parent" to the FIPE3 question. If the FIPE3 question was answered "yes," the RELMATCH variable had a value of 3, and the indirect count was selected as the multiplicity count. If the FIPE3 question was answered "no," the pair was not considered a child-parent pair, and would not be considered for these counts. Finally, if the FIPE3 question was not answered, the respondent was considered a "parent" if he or she was a stepparent. If the respondent was a live-in partner, the determination of the pair relationship was left to imputation. The multiplicity count was set to the indirect count to account for the possibility that the pair relationship would be imputed as parent-child.

2. Suppose the selected child did not accept a stepparent or live-in partner as his/her parent (as above), but the other respondent selected was the "true" or "original" parent. In this case, the stepparent or live-in partner was only identified in the "original" parent's roster, so there was no way to determine how the stepparent or live-in partner would have answered the FIPE3 question. The stepparent was considered a "parent" even if the child did not view him or her this way, so that the indirect count was used. The case of live-in partners was less clear. If the live-in partner had been selected, the determination of whether a parent-child relationship was indicated would have involved the response to the FIPE3 question, which we didn't have since the live-in partner was not selected. Hence, these cases were left to imputation.

Parent-child pairs, parent focus. The multiplicity counts in this domain reflected the selected parent's children, and were limited to have values of at least 1. If neither side had bad relationship codes, the following rules applied:

1. In most cases, if one pair member had bad data, the multiplicity was obtained from the other pair member. The exception was when the number of household members between 12 and 14, 12 and 17, 12 and 20, or 15 and 17, (depending on the domain) in the "bad side" matched the number in the corresponding age ranges in the screener roster, but the "good side" had a larger number in the corresponding age ranges than in the screener roster, a larger number which matched the multiplicity count of the "good side." The larger number was due to the fact that the "good side" originally had no "self" identified in its roster, and an extra roster member was incorrectly added to the "good side" in the roster editing stage to create a "self" (see Section 6.2.2.2 in the main body of the report). In this case, the count of individuals within the age range on the "bad side" was used as the final count. (This was only an issue in the 1999 survey year.)

2. If the count of children in the household within the relevant age ranges differed between the pair members, but one side had a count of children equal to the same count from the screener roster, the multiplicity count that corresponded to the pair member with the same count of children as the screener was used.

3. If the count of children in the household within the relevant age ranges differed between the pair members, and both sides had a multiplicity count that exceeded the count of all children from the screener roster, the number of children in the screener roster was used as the multiplicity count. If the screener roster had missing exact ages, then the minimum multiplicity count from the two pair members' rosters was used as the final count.

4. The direct count and indirect count might differ because either the child lists a sibling that the parent considers "another relative," or the parent lists a child that the child considers "another relative." In either case, since the parent was the one to answer the FIPE3 question, the multiplicity count from the parent's perspective was selected as the final count.

R.3 Sibling-sibling counts

Although there were two types of sibling-sibling pairs under consideration, each associated with two domains, the same rules could be applied to all four domains. When the older sibling was the focus, the multiplicity count was a count of the number of siblings within the younger age group (12 to 14 or 12 to 17). The younger age ranges in these rules can be switched to the older age ranges when the younger sibling was the focus. The following general rules apply:

1. The counts disagreed if a household member left or entered the household between interviews. As before, the roster that was closest to the screener was used to determine the count. If one roster member had the same number of household members within the ages of 12 to 14 or 12 to 17 (depending on the domain) as the screener roster, the multiplicity count from that roster member was used, provided the member had no bad relationship codes within the relevant age range.

2. If the counts disagreed and both exceeded the screener count of household members within the relevant age range, the multiplicity count was set to the screener count. If the screener roster had missing exact ages, then the minimum multiplicity count from the two pair members' rosters was used as the final count.

3. If the younger pair member identified the older as "sibling" but the older pair member did not reciprocate, then imputation was required to establish whether the relationship was sibling-sibling. For those pairs that were imputed to sibling-sibling, the count was incremented by 1 to reflect the fact that the younger sibling's relationship code was changed from nonsibling to sibling. However, if the younger sibling identifies other siblings within the relevant age range that the older sibling did not identify, then it was necessary to accept the direction of the imputation–that is, to identify these other roster members as siblings.

4. The counts disagreed if the siblings disagreed on whether one or more household members within the relevant age range was a sibling of theirs. However, if the minimum number of respondent's children possible, considering age ranges and sibling codes within both questionnaire rosters and the screener, was equal to the maximum number possible, then the counts were set to the equal bounds. Otherwise, there was no way to reconcile these differing counts, so the final count was left to imputation, within the bounds determined by the two pair members' counts.

5. Other counts that were left to imputation involved cases where both sides had too many bad relationship codes to definitively determine a multiplicity count.

Appendix S

Conditions Used for Reconciling Differing Household-Level Person Counts between Pair Members

Appendix S

Conditions Used for Reconciling Differing Household-Level Person Counts between Pair Members

S.1 Introduction

Household-level person counts for a particular domain were obtainable using the multiplicity counts if the pair belonged to a pair relationship that fit into that domain, provided only one family unit was in the household. No reconciliation between pair members was necessary in that case, since the reconciliation had already been done with the multiplicity counts. Other counts were obtained from single respondents, for whom no reconciliation was necessary. This appendix discusses the conditions used to reconcile differing household-level person counts when the pair belonged to a pair relationship that corresponded to different pair domains than the one being counted. Typically, the counts between the two pair members were in agreement, and the determination of the final household-level count was straightforward, provided both rosters did not have bad data codes.[23] The strategy was also usually clear if one pair member had bad data in the household roster; the count from the pair member with good data was usually preferred in those cases. If the bad data was limited to bad relationship codes, then the member with good data was only selected if substituting the appropriate relationship codes for the bad data codes would have given a total that was equal to the count from the pair member with good data. There were instances where bad data codes existed in the roster, and this condition did not apply. There were other exceptions as well. Finally, there were instances where neither pair member had bad data in their rosters, yet their counts still disagreed. In this appendix, the rules that were used to reconcile these disagreeing counts are outlined. For each pair domain, a set of general rules are given, each with specific conditions required for the general rule to be implemented. Within each general condition, if at least one of the specific conditions was not satisfied, upper and lower bounds were determined and the final count was left to imputation.

S.2 Parent-child counts

For parent-child counts where the pairs were not parent-child pairs of interest (e.g., sibling-sibling pairs, parent-child pairs where the child was 21 or over, etc.), the screener was used to help reconcile disagreeing counts. The rules follow below, separated by the member of focus:

[23] If a roster pointed to a household size of 1, this was considered "bad data," since both pair members in the household were survey respondents.

Parent-child pairs, child focus. For the child-focus counts, the count is of the number of children of a parent in the household. The following general rules applied:

1. Among non-parent-child pairs of interest, in most cases, the counts for the two sides agreed. However, both sides had to meet the following conditions, in order for the final count to be set to one of the sides:

    - Either no bad ages with the relevant relationship codes and no bad relationship codes within the relevant age ranges, or the counts were equal to the screener age counts, or a side with good data indicated siblings within the relevant age range living together in a household without parents.
    - No situations where parents were not identified in the household, but some in the household had bad relationship codes and were old enough to be parents.
    - No counts of one child in the relevant child-age range when both members of the pair were in that range, and the children were siblings.
    - No pairs where the ages of the identified parents did not match and both sides had relationship codes indicating "other relative" or a nonrelative, indicating more than one family unit in the household.[24]
    - The household size was greater than 1 and non-missing on both sides.

2. The counts might have agreed even though the above conditions were not met. The count could still be set to one of the sides, if any one of the following was true:

    - If the number of children matched across both rosters and the screener.
    - If the counts which agreed with each other equaled or exceeded the count of the number of children from the screener.
    - If both sides had a count of 0, both had a roster, and (at least) one side had all good age and relationship codes.
    - If both sides had a count of 0, both had a roster, and the number of respondents who were old enough to be parents in the household was 0 according to the screener.

3. The counts might have agreed with a value of 1. If both pair members were children within the relevant age range, and both indicated they had parents even though the children were siblings, then they were not included in each other's rosters, but were obviously in the screener roster, so the count was set to 2.

4. If one pair member did not have a valid roster but the other member did, the final count was set to the other pair member's count under the following conditions:

    - Either:

[24] Codes which indicate "other relative" or a non-relative are 7 (roommate), 8 (child-in-law), 10 (parent-in-law), 12 (boarder), 13 (other relative), and 14 (other non-relative).

- There were no bad relationship codes within the relevant child-age ranges and the respondent identified parents in the household, or
- There were no children within the relevant age range, or
- No parents were identified in the household and nobody in the roster older than the respondent had a bad relationship code.

- No counts of one child were in the relevant child-age range when both members of the pair were in that range, and the children were siblings.

5. If one pair member did not have a valid roster but the other member did, and the above conditions were not met, it was still possible to use the other pair member's count, if that count was zero, under any of the following conditions. Either

   - The other roster was valid, did not have any bad ages, and had no ages in the relevant age range, or
   - The other roster was also bad, but the screener roster was valid, and did not have any ages in the relevant age range, or
   - The respondent identified both grandchildren and grandparents in the roster, where the grandchildren referred to the grandparent instead of the respondent.

6. When two different family units were in the household, the determination of the final count had to be treated separately. This could have included the multi-generational families referred to earlier, and to two siblings both with children in the relevant age range living in the household. The latter was more easily identified if it was not a parent-child pair (e.g., a cousin-cousin pair). The sum of the two counts (one might be zero) was used provided the following conditions were satisfied on both sides:

   - There were no bad ages or relationship codes within the relevant age ranges.
   - Both had counts pointing to 2 or fewer parents, meaning that the two family units were not identifiable on a side.
   - The number of identified parents were not equal to the total number over 25 in the household on either side, meaning that parents could correspond to roster members identified by other relationship codes.
   - The number of identified children were not equal to the total number within the relevant age range in the household on either side, meaning that children with parents could correspond to roster members identified by other relationship codes.
   - There were not three generations in the household, with first and second generation parents both having children in the appropriate age range. This was already accounted for by the counts for one or both sides.

If the above conditions were not met, the two families in the household might have been already accounted for when the counts were determined for each side. The maximum of the two counts was taken if the household members in the roster over 25 (of parental age) were either both equal to the number over 25 in the screener roster, or both different than the number over 25 in the screener roster. However, if the number over 25 in the screener

roster was equal to the number over 25 in one of the pair member's rosters, but not the other, then the count where the number was equal to the screener roster was taken. In both instances, the count corresponding to the pair member that matched the screener roster was taken.

7. If one pair member did not have a valid roster, and the pair member with a valid roster was within the valid age range and was a sibling to the other pair member, but the count from his roster was only 1, the count was set to 2.

8. If the pair relationship was not parent-child nor was it sibling-sibling, but one side had nonzero counts and the other did not, it was necessary to decide who to believe. Often this occurred because one of the respondents was a relative outside the nuclear family unit, like a cousin or aunt/uncle, whose own parents did not live in the household, or a boarder.[25] Selecting either the zero count or nonzero count in this instance required that:

   - The respondent with zero count did not identify parents in the roster or he/she identifies parents but was over 20 years old, and had no bad relationship codes within the relevant age ranges.
   - Either the respondent with nonzero count had siblings or children within the relevant age range, or was himself/herself within that age range (with a count of 1).

   When one count was zero and the other nonzero, the nonzero count was used under the following conditions

   - The respondent pair member with nonzero count did not have bad relationship codes,
   - Either:
     - The count of children within the relevant age range in the household for the nonzero count pair member matched that of the zero count pair member, or
     - The count of children in the household within the relevant age range for the nonzero count pair member matched that of the screener, or
     - The count of children in the household within the relevant age range for the zero count pair member matched that of the screener, because a child was (or children were) listed as 11 years old in the nonzero count pair member's roster, when he or she (they) should have been 12 (according to the nonzero count pair member's and the screener roster) so that the final count was the nonzero count with this child (these children) added, or
     - The respondent with zero count had no household members with a family-type relationship code, or

[25] Even if there was disagreement between the respondents about whether a boarder or other family member was in fact a sibling, parent, or child, this would had been resolved at the pair relationship stage, where we would had determined whether this was in a domain of interest.

- The count of children within the relevant age range in the household for the zero count was closer to the screener age count, but the nonzero count, was less or equal to than the screener age count, or
- The other conditions had not already established a nonzero count, but a count for a subset age group had already been established as nonzero. For example, if the count for 12 to 14 year olds was nonzero, then the 12 to 17 year old count had to be nonzero.

The zero count was used if:

- The household age composition among the relevant age ranges for the zero count pair member more closely matched the screener, or
- The pair was a grandparent-grandchild pair with an adult child of the grandparent living in the household. The nonzero count resulted from an assumption that a respondent's adult child and grandchild within the relevant age range were a parent-child pair. If the grandchild identified the grandparent's child as "other relative," and did not identify any parents, this indicated that the grandparent's adult child was an uncle/aunt of the grandchild, not a parent.

9. Even with sibling-sibling and parent-child pairs, sometimes one side had a zero count and the other had a nonzero count. This was usually due to one pair member having missing relationship codes for the roster member that would have been identified as a parent (i.e., relationship codes for roster members in a parental age range). If the count for the pair member with the entirely good roster was equal to the number within the appropriate age range for the pair member with bad relationship codes in the roster, the nonzero count was selected.

10. The two counts might have disagreed because one side had bad relationship codes within the relevant age range, and the other did not. If the sum of the number of bad relationship codes with the smaller count equaled the larger count, the larger count was chosen.

11. The two counts might have disagreed because they disagreed on the ages of one or more household member, even though each respondent's count included all the children in their respective roster. If the roster for one respondent more closely matched the screener in terms of the distribution of ages within the roster, then that respondent's count was chosen. If the screener roster was a valid roster, but had fewer children in the relevant age range than the nonzero count of either pair member, then the final count was set to the number of children in the relevant age range of the screener roster.

12. The two counts might have disagreed because they disagreed on the ages of one or more household member, and each respondent's count included all the children in their respective roster, but neither was closer to the screener count. If the screener count differed from each respondent's count by the same amount, was greater than one but less than the other, the screener count was used as the final count.

13. If the pair relationship was parent-child, and the parent-child counts were associated with the same age range, then the household-level person counts would have been obtained using the parent-focus multiplicity counts. However, this did not occur if the age range for the pair relationship differed from the age range for the parent-child counts. If the pair relationship was imputed to be parent-child, or it was deemed parent-child even though the child did not consider the parent a "parent," but the parent answered the FIPE3 question, the nonzero count should be used as the final count.

14. If after all the above tests were done to find the final count, the minimum possible and maximum possible counts, considering both questionnaire rosters and the screener roster, were the same, then the final count was set to that value.

15. Remaining disagreeing counts were left to imputation, with appropriate bounds set on the imputed value.

Parent-child pairs, parent focus. For the parent-focus counts, the count is of the number of parents of at least one child in the household The child-focus parent-child counts are processed first, so if the child-focus parent-child counts are zero, it necessarily means that the parent-focus counts will also be zero. Nonzero child-focus counts also point to nonzero parent-focus counts. After setting counts to 0 where necessary, the following general rules applied:

1. Among non-parent-child pairs of interest, in most cases, the counts for the two sides agreed. However, both sides had to meet the following conditions, in order for the final count to be set to one of the sides.

   - No situations where both pair members were children in the relevant age range, but were in a spouse-spouse pair relationship, and both identified the same roster member as parent,
   - Either:
     1. No bad relationship codes for household members of an age to be parents, or
     2. The total count was 2, for 2 parents, or
     3. The total count + the number of grandparents equaled the total number 26 or over in the household, according to the screener roster.
   - The household size was greater than 1 and non-missing on both sides.

   Note that it was not necessary to check for bad relationship codes in the child age ranges, since it was already known that the count had to be at least 1, and the number of children was not important for the parent counts.

2. The counts may have agreed even though the above conditions were not met. The final count could still have been set to one of the sides if it was a sibling-sibling pair, and the bad codes in the parental age range were on one side only. This would indicate that the side with bad codes were not missing parental codes.

3. If one pair member did not have a valid roster but the other member did, the final count was set to the other pair member's count if there were no bad relationship codes and no roster members with bad age and bad gender values. Other circumstances called for setting the final count to zero, which would necessarily be the case if the child-focus counts were zero.

4. When two different family units were in the household, the determination of the final count had to be treated separately. This could have included multi-generational families, or two siblings both with children in the relevant age range living in the household. The latter was more easily identified if it was not a parent-child pair (e.g., a cousin-cousin pair). The sum of the two counts (one might be zero) was used under the following conditions:

   - There were no bad ages or relationship codes within the relevant age ranges.
   - Both had counts pointing to 2 or fewer parents, meaning that the two family units were not identifiable on a side.
   - The number of identified parents were not equal to the total number over 25 in the household on either side, meaning that parents could correspond to roster members identified by other relationship codes.
   - There were not three generations in the household, with first and second generation parents both having children in the appropriate age range. This was already accounted for by the counts for one or both sides.

   If the above conditions were not met, the two families in the household might have been already accounted for when the counts were determined for each side. The maximum of the two counts was taken if the household members in the roster over 25 (of parental age) were either both equal to the number over 25 in the screener roster, or both different than the number over 25 in the screener roster. However, if the number over 25 in the screener roster was equal to the number over 25 in one of the pair member's rosters, but not the other, then the count where the number was equal to the screener roster was taken. In both instances, the count corresponding to the pair member that matched the screener roster was taken.

5. If the pair relationship was a spouse-spouse pair, and one of the pair members had a positive count, with an age within the relevant child age-range, then the count for that pair member was taken as the final count, provided there were no bad relationship codes in that roster for roster members aged 18 or over.[26]

6. The two counts might have disagreed with one count nonzero, and the other equal to 0. In order to make it to these conditions, the count had to be nonzero. The nonzero was chosen as the final count if:

[26] In almost all cases (all cases in the 2002 survey year), either the count for the other pair member was 0, or the count for the pair members was equal. In the latter case, one of the identified parents should have been "parent-in-law." In one case in the 2001 survey year, the counts were both nonzero and unequal. In that instance, this condition should not have been invoked. The software has been corrected for the 2003 survey year.

- The count was 1, and there were no bad ages with the relevant relationship codes and no bad relationship codes within the relevant age ranges, or
- The count was 2.

7. The two counts might have otherwise disagreed, where the number of roster members 26 or over disagreed between the two pair members. In these situations, one count was 1, and the other 2. The final count corresponded to the pair member with the number of roster members 26 or over closest to the screener number of roster members 26 or over, under the following conditions:

   - The difference between the screener count of the number of household members 26 or over, and the pair members' counts of this number of household members was not the same between the two pair members.
   - Neither pair member had bad ages in their rosters.
   - Each pair member either had no bad relationship codes in his or her roster, or had a nonzero count with no bad relationship codes among respondents 26 or over.

8. The two counts might have otherwise disagreed if the bad relationship codes referred to missing parental codes. If one side had no bad relationship codes, then the sum of the number of bad relationship codes and the count on the side with the bad codes was equal to the count on the side with no bad relationship codes.

9. The two counts might have disagreed where one count was 2, and the other was 3. Since households with two family units had already been considered, the maximum number of parents possible was 2, so the final count was set to 2.

10. If after all the above tests were done to find the final count, the minimum possible and maximum possible counts, considering both questionnaire rosters and the screener roster, were the same, then the final count was set to that value.

11. Remaining disagreeing counts were left to imputation, with appropriate bounds set on the imputed value.

S.3 Sibling-sibling counts

The logic for the sibling-sibling counts did not depend upon whether the younger age range was 12 to 14 or 12 to 17, or whether the older age range was 15 to 17 or 18 to 25. It also did not depend upon which pair member was the focus, though for the household-level person counts, the older member focus counts were the only ones considered. Hence, the counts that are of interest are of roster members in the older age range. As with the parent-child pairs, the multiplicity counts could be used if the pair relationship was a sibling-sibling pair of interest. However, the counts had to be determined for all other pairs. The rules follow below, separated by the member of focus:

1. Among pairs that were not sibling-sibling pairs of interest, in most cases, the counts for the two sides agreed. However, both sides had to meet the following conditions, in order for the final count to be set to one of the sides:

   - The pair could not be a sibling-sibling pair, where both respondents were in the older age range, and have a younger sibling in the younger age range, and the count was 1. (This refers to a sibling-sibling pair that would not constitute a domain of interest.)
   - No bad relationship codes in the lower range if the count was 0.
   - Either:
     - No bad relationship codes in the upper range, or
     - The count matched the screener age count.
   - The household size was greater than 1 and nonmissing on both sides.

2. The counts might have agreed even though the above conditions were not met. The count could still be set to one of the sides, if any one of the following was true:

   - If the number of children matched across both rosters and the screener, for both the upper and lower age ranges.
   - If the count was 0, and one of the two was true:
     - Neither side had bad relationship codes or ages, or
     - The number of household members in the screener 26 years of age or older was 0.

3. If one pair member did not have a valid roster but the other member did, the final count was set to the other pair member's count under the following conditions:

   - No bad relationship codes within the lower age range when the count was zero.
   - Either:
     - There were no bad relationship codes within the upper age range, or
     - The count was equal to the screener age count within the upper age range, or
     - The count was zero, and the count of household members in the lower age range was zero.

4. If one pair member did not have a valid roster but the other member did, and the above conditions were not met, it was still possible to use the other pair member's count, under the following conditions:

   - The count was zero,
   - Either:
     - The number of children in either the lower or upper age ranges was zero with no bad ages in the roster, or

- The number of children in the screener roster in either the lower or upper age ranges was zero, with a valid screener roster.

5. When two different sets of siblings were in the household, the determination of the final count had to be treated separately. The two sets of siblings refer to siblings where both parents from one set differ from the parents of the other set. The sum of the two counts (one might be zero) was used, provided the following conditions were satisfied for both pair members:

   - There were no bad relationship codes within the upper age ranges.
   - There were no bad relationship codes within the lower age range, or the count was nonzero.

6. If the counts from the two pair members did not agree, the following rules were used to assign the appropriate count, provided no bad relationship codes were evident in either age range, on either side. These conditions are hierarchical, in that subsequent conditions require that the previous condition was not met.

   - If the number within the upper age range was the same on both sides, but the number in the lower age range was not, the side was chosen with the number in the lower age range equal to the number in the screener roster within the lower age range. (In all cases, one size had zero count and the other did not. This captured situations where it was necessary to discern whether the zero count was due to no children in the lower age range on one side, and whether the screener also had no children in that range.)
   - For one pair member, the number of children in either the lower age range or the upper age range did not agree with the number in the screener in that range. However, for the other pair member, the number within both age ranges agreed with the screener count. The count was set to the side that agreed with the screener.
   - For both pair members, the numbers within the lower age range were either both zero, or both positive. The number within the upper age range did not agree between pair members, but one pair member agreed with the screener. The count was set to the count for that pair member.
   - In the rosters for both pair members and the screener, the numbers within the upper age range nonzero for at least one of the three were nonzero, but not necessarily equal. The numbers within the lower age range were not equal across any of the three rosters. The pair member with the number of children in the younger age range closest to the screener was selected.
   - In the rosters for both pair members and the screener, the numbers within the lower age range nonzero for at least one of the three were nonzero, but not necessarily equal. The numbers within the upper age range were not equal across any of the three rosters. The pair member with the number of children in the upper age range closest to the screener was selected.

7. If the counts from the two pair members did not agree, but one side had bad relationship codes within the upper age range, and the other did not have bad relationship codes, and the sum of the count and the number of bad relationship codes on one side was equal to the count for the pair member with the good roster, the count for the pair member with the good roster was selected.

8. If the counts from the two pair members did not agree, but the above conditions were not met, in many cases this was due to one of the pair members not being part of the immediate family unit, in which case his or her count was automatically zero. To identify these cases, and assign the count to the other pair member, the following conditions had to be satisfied:

   - The pair relationship did not indicate an identifiable family-type relationship (e.g., sibling-sibling, parent-child, spouse-spouse, or grandparent-grandchild relationship).
   - Either:
     - One pair member did not have any relationship codes indicating parent, child, sibling, spouse, grandchild, or grandparent, and
     - The other pair member had at least one relationship code indicating a relationship other than parent, child, sibling, spouse, grandchild, or grandparent, and
     - For the pair member with family codes, either no bad relationship codes were within both the upper and lower age ranges, or no bad relationship codes were within the upper age range, and the count was positive,

     or

     - There were no bad relationship codes within both the upper and lower age ranges for either pair member.

9. If one pair member had no bad relationship codes within both the upper and lower age ranges, but the other had some bad codes, then the count associated with the pair member with no bad codes was selected if the count of immediate family members (parent, child, sibling, spouse, grandchild, grandparent) was the same as the count of household members within both the lower and upper age ranges.

10. If one pair member had a zero count due to having no household members within the upper age range, but the number of household members within that age range was nonzero for both the screener and the other pair member (though not necessarily equal), then a nonzero count was selected. If the count for the other pair member was equal to the number of household members within the upper age range for that pair member, then the final count was set to the screener number of household members within that age range.

11. If the pair was a spouse-spouse pair, one count might have been zero while the other was nonzero because the spouse-spouse pair still lived with the parents of one

pair member, and the pair member's younger siblings also lived in the household. In this case, the nonzero count was selected if the number of immediate family members (parent, child, sibling, spouse, grandchild, grandparent) if the roster for the pair member with the zero count was less than his or her total household size.

12. In some cases, one pair member called the other pair member a parent or child, but the other pair member did not reciprocate. In the case of a child who did not reciprocate the parent's identification of him or her as child, the child's count was always less than the parent's count. By the same token, in the case of a parent who did not reciprocate the child's identification of him or her as parent, the parent's count was always less than the child's count. If the pair relationship was imputed to be "parent-child," then the pair member who did not acknowledge a parent-child relationship was overruled, and the maximum count of the two pair members was selected as final.

S.3 Spouse-spouse counts (with or without children)

The multiplicity counts were not useful in the logic for the spouse-spouse household counts, since the spouse-spouse multiplicity counts were always 1. The logic for the spouse-spouse counts follows:

1. Among the majority of pairs, the counts for the two sides agreed. However, both sides had to meet the following conditions, in order for the final count to be set to one of the sides.

   - The pair could not be a spouse-spouse pair, where both respondents had a spouse or both respondents had a partner,
   - No bad relationship codes for roster members 15 or over,
   - The number of spouse-spouse pairs was either 1 or 0 for both pair members,
   - There was no potential for 2 or more couples in the household that were not already obviously identified, whereby one of the pair members had grandchildren and there were respondents 12 or over who were not children, grandchildren, siblings, children, parents, spouses, or partners, but no children-in-law were identified, and
   - The household size was greater than 1 and nonmissing on both sides.

   NOTE: This general condition failed to exclude some of the cases where one couple was identified by both pair members, but the identified couple was different for each pair member. This occurred most commonly with multgenerational families with two couples in the household, where the spouse/partner in the younger couple who "married into" the family did not recognize the spouse/partner's parents as parents-in-law. This has been corrected for processing in years subsequent to 2002.

2. The counts might have agreed even though the above conditions were not met. The count could still be set to one of the sides, if any one of the following was true:

- There was no potential for 2 or more couples in the household that were not already obviously identified, whereby one of the pair members had grandchildren and there were respondents 12 or over who were not children, grandchildren, siblings, children, parents, spouses, or partners, but no children-in-law were identified.
- Either:
  - One pair member has a single bad relationship code, and no other relationship codes could match it to make it a couple (i.e., the pair member does not have a single identified parent, grandparent, parent-in-law, or child-in-law). The other pair member has no bad relationship codes.
  - One pair member has bad relationship codes among roster members 15 or over, or has bad ages, and the other has no bad ages or relationship codes, where the pair member with no bad roster entries has the same age composition as the screener. The pair member with the bad roster entries would have the same age composition as the screener if the number of roster members 15 or over was added to the number of roster members with bad ages.
  - One pair member has bad relationship codes among roster members 15 or over, or has bad ages, and the other has no bad ages or relationship codes, where all the relationship codes for the pair member with no bad roster entries are immediate family codes (child, parent, sibling, spouse, partner, grandparent, or grandchild). For the pair member with bad roster entries, all the existing relationship codes are immediate family codes.

3. If the household size was 1, or the number of respondents 15 or over in the household was 1 or 0, then the count should automatically be zero. Instead of setting the count to zero, the code set the count to pair member A's count. In a very small number of cases (2 cases in the 2002 survey), this count was 1 instead of 0, which was an error.

4. For those cases where the pair was imputed to be a spouse-spouse pair, and both sides agreed that only one spouse-spouse pair was in the household, the count was set to one if:

   - There was no potential for 2 or more couples in the household that were not already obviously identified, whereby one of the pair members had grandchildren and there were respondents 12 or over who were not children, grandchildren, siblings, children, parents, spouses, or partners, but no children-in-law were identified, and
   - Either:
     - Both sides had fewer than 4 people older than 15 in the household, or
     - One side had fewer than 4 people older than 15 in the household, and the other had no bad relationship codes among roster members 15 or over.

5. If one pair member did not have a valid roster but the other member did, the final count was set to the other pair member's count under the following conditions:

- There was no potential for 2 or more couples in the household that were not already obviously identified, whereby one of the pair members had grandchildren and there were respondents 12 or over who were not children, grandchildren, siblings, children, parents, spouses, or partners, but no children-in-law were identified, and
- Either:
  - There were no bad relationship codes among roster members 15 or over, or
  - There were no bad relationship codes among roster members 18 or over, and the pair member had parents.

6. When two different family units were already identified in the household, then two different parent-sets were being referenced (one of the parent-sets was often a single parent). The sum of the two counts (one might be zero) was used provided neither pair member had grandparents or grandchildren identified. This was to prevent spouse-spouse pairs from being counted twice, which would happen if grandparents were also parents of 0 to 17 year olds. If two family units were multigenerational families, then the final count was obtained by taking the maximum of the two pair members' counts.

7. It was possible for two different spouse-spouse pairs to be in the household, even though two different family units had not been identified. The final count was set to 2, even though two family units had not been previously identified, under the following conditions:

   - The pair relationship was not a spouse-spouse pair, and the total household size was at least 4, and
   - Either:
     - Both sides identified a spouse, or
     - Both sides identified a partner, or
     - One side identified a parent and the other identified a parent-in-law.

8. If the count of the number of spouse-spouse pairs did not agree between the two pair members, it could be because a couple entered the household or otherwise materialized after screening. The smaller count was chosen as the final count in this instance, which was identified if the following conditions were satisfied:

   - There was no potential for 2 or more couples in the household that were not already obviously identified, whereby one of the pair members had grandchildren and there were respondents 12 or over who were not children, grandchildren, siblings, children, parents, spouses, or partners, but no children-in-law were identified.
   - The screener count of roster members 12 or over was no larger than the count of roster members 12 or over in the roster of the pair member with the smaller spouse-spouse count.

- The screener count of roster members 12 or over was smaller than the count of roster members 12 or over in the roster of the pair member with the larger spouse-spouse count.
- The difference between the screener count of roster members 12 or over and the count of roster members 12 or over in the questionnaire rosters of the pair members was smallest with the pair member with the smaller spouse-spouse count.

9. If the count of the number of spouse-spouse pairs did not agree between the two pair members, it could be because a couple left the household or otherwise dissolved after screening. The larger count was chosen as the final count in this instance, which was identified if the following conditions were satisfied:

    - There was no potential for 2 or more couples in the household that were not already obviously identified, whereby one of the pair members had grandchildren and there were respondents 12 or over who were not children, grandchildren, siblings, children, parents, spouses, or partners, but no children-in-law were identified.
    - The screener count of roster members 12 or over was no larger than the count of roster members 12 or over in the roster of the pair member with the larger spouse-spouse count.
    - The screener count of roster members 12 or over was larger than the count of roster members 12 or over in the roster of the pair member with the smaller spouse-spouse count.

10. In many cases where the count of the number of spouse-spouse pairs did not agree between the two pair members, one side had zero count and the other did not. The nonzero count was selected if the pair member associated with the zero count was not a close relative, or somehow otherwise did not identify a spouse, partner, 2 parents, or 2 grandparents. The following conditions were required to select the nonzero count:

    - There was no potential for 2 or more couples in the household that were not already obviously identified, whereby one of the pair members had grandchildren and there were respondents 12 or over who were not children, grandchildren, siblings, children, parents, spouses, or partners, but no children-in-law were identified.
    - The pair member with a nonzero count either identified a spouse, a partner, two parents, or two grandparents.
    - The number of roster members 15 or over associated with the nonzero count pair member was no larger than the corresponding number associated with the zero count pair member.
    - If the side associated with the nonzero count identified a spouse, partner, or 2 parents, the following additional conditions were required:

- The number of roster members between 26 and 44 was the same between the two pair members.
- The number of roster members between 30 and 49 was the same between the two pair members.
- The number of roster members between 35 and 54 was the same between the two pair members.
- The number of roster members between 40 and 59 was the same between the two pair members.

- If the side associated with the nonzero count identified 2 grandparents, the following additional condition was required:
  - The number of roster members 50 or over was the same between the two pair members.

11. If either a pair member's partner was not considered a family member by the other pair member, or if a pair member had two grandparents, and an uncle/aunt husband-wife pair in the household, then the maximum was selected under the following conditions:

    - There was no potential for 2 or more couples in the household that were not already obviously identified, whereby one of the pair members had grandchildren and there were respondents 12 or over who were not children, grandchildren, siblings, children, parents, spouses, or partners, but no children-in-law were identified.
    - Either:
      1. At least one side identified a partner, and the maximum count was 1, or
      2. The pair member associated with the smaller count had a grandparent, and had at least 2 roster members who were not either a parents, siblings, children, spouses, partners, or grandparents.

    Note: this condition did not consider cases where the difference in counts was due to different household compositions between the pair members.

12. The count of the number of spouse-spouse pairs might not agree because one of the pairs was a sibling and sibling-in-law, and there are no codes for sibling-in-law. The maximum count was selected under the following conditions:

    - There was no potential for 2 or more couples in the household that were not already obviously identified, whereby one of the pair members had grandchildren and there were respondents 12 or over who were not children, grandchildren, siblings, children, parents, spouses, or partners, but no children-in-law were identified.
    - The pair member with the smaller count did not have a spouse or partner, but did have siblings aged 15 or over, and there were household members in his or

her roster that were not parents, children, siblings, spouses, partners, grandchildren, or grandparents.

13. The count of the number of spouse-spouse pairs might not agree because one side had no nuclear family or grandparent-grandchild relationship codes, and one of the selected respondents was not in a child-parent, child-grandparent, or spouse-spouse relationship. The maximum count was selected if:

    - There was no potential for 2 or more couples in the household that were not already obviously identified, whereby one of the pair members had grandchildren and there were respondents 12 or over who were not children, grandchildren, siblings, children, parents, spouses, or partners, but no children-in-law were identified,
    - The pair member's roster associated with the minimum count (usually zero) had no children, parents, siblings, spouses, partners, grandchildren, or grandparents among respondents 12 or over, and
    - The pair member's roster associated with the maximum count had some roster members who weren't children, parents, siblings, spouses, partners, grandchildren, or grandparents.

    Note: this condition also nabbed cases where the relationship codes were not correctly identified on one pair member's roster. This occurred rarely, but when it did, the minimum count was 1 and the maximum count was 2.

14. The count of the number of spouse-spouse pairs might not agree because the pair were siblings, but one sibling did not consider a step-parent or parent's partner as a "parent." The maximum count was selected if:

    - There was no potential for 2 or more couples in the household that were not already obviously identified, whereby one of the pair members had grandchildren and there were respondents 12 or over who were not children, grandchildren, siblings, children, parents, spouses, or partners, but no children-in-law were identified,
    - The pair members were siblings,
    - The pair member associated with the maximum count had two parents,
    - The pair member associated with the minimum count had one parent, and
    - The roster associated with the pair member with the maximum count had more immediate family members (children, parents, siblings, spouses, partners, grandchildren, or grandparents) than the roster associated with the other pair member.

15. The count of the number of spouse-spouse pairs might not agree because the household otherwise changed after screening, which was not accounted for by previous conditions. In general, the count with a household composition closest to the screener was selected. The age composition was defined by looking at age

classes. The count for a given pair member was selected if the following properties held:

- There was no potential for 2 or more couples in the household that were not already obviously identified, whereby one of the pair members had grandchildren and there were respondents 12 or over who were not children, grandchildren, siblings, children, parents, spouses, or partners, but no children-in-law were identified.
- The number of roster members between the ages of 26 and 44 for that pair member matched the screener count within the same age range, which differed from the corresponding count for the other pair member.
- The number of roster members between the ages of 30 and 49 for that pair member matched the screener count within the same age range, which differed from the corresponding count for the other pair member.
- The number of roster members between the ages of 35 and 54 for that pair member matched the screener count within the same age range, which differed from the corresponding count for the other pair member.
- The number of roster members between the ages of 40 and 59 for that pair member matched the screener count within the same age range, which differed from the corresponding count for the other pair member.

16. In some cases, neither pair member's household composition matched that of the screener. In that case, the household roster closest to that of the screener was selected. The maximum was selected if the following conditions were satisfied:

    - There was no potential for 2 or more couples in the household that were not already obviously identified, whereby one of the pair members had grandchildren and there were respondents 12 or over who were not children, grandchildren, siblings, children, parents, spouses, or partners, but no children-in-law were identified.
    - The number of screener roster members aged 12 or over exceeded the corresponding count from the questionnaire rosters of both pair members, which also differed from each other.

17. If the counts did not match, on the rare occasion one pair member in a spouse-spouse pair identified two grandparents of a different gender. Since there is no code for grandparents-in-law, they could not be identified, so the maximum count was selected. The following conditions were required:

    - There was no potential for 2 or more couples in the household that were not already obviously identified, whereby one of the pair members had grandchildren and there were respondents 12 or over who were not children, grandchildren, siblings, children, parents, spouses, or partners, but no children-in-law were identified.
    - The pair was a spouse-spouse pair.

- The pair member with the maximum count had 2 grandparents of a different gender, and the pair member with the minimum count did not have any.

The assumption here, of course, is that the grandparents of a different gender are in fact a spouse-spouse pair. There is no way to check whether a grandfather is the father's father, and the grandmother is the mother's mother, for example.

18. Even though the household composition may match in terms of ages across the screener roster and the two pair members' rosters, the counts may disagree where two spouse-spouse pairs were clearly identified by one pair member but not the other. This may be because one of the in-laws was incorrectly identified on one side, or because a partner was not considered an in-law by a responding pair member, or because a partner did not consider other family members as "in-laws." The following conditions were required for the maximum count to be selected:

    - There was no potential for 2 or more couples in the household that were not already obviously identified, whereby one of the pair members had grandchildren and there were respondents 12 or over who were not children, grandchildren, siblings, children, parents, spouses, or partners, but no children-in-law were identified.
    - The number of screener roster members aged 12 or over matched the corresponding count from the questionnaire rosters of both pair members.
    - The pair member with the maximum number of spouse-spouse pairs had a spouse or partner, and also had two parents.
    - There were no bad relationship codes among roster members 15 or over on either pair member's roster.

19. If the counts for each pair member are not equal, but the number of roster members aged 12 or over is the same between the two pair members, and the count for one pair member is the maximum possible in the household, then that number is selected as the final count. This condition is only applied after all other conditions, including conditions where the final count is ambiguous, have already been applied.

S.4 Spouse-spouse counts (with children)

The household counts for spouse-spouse counts with children obviously depended upon the counts obtained for spouse-spouse counts with or without children. The logic for the spouse-spouse counts with children follows:

1. For a sizable proportion of cases, clearly no couples with children could be in the household, either because the spouse-spouse count was zero, or because the household size was 2 or less. In these cases, the final spouse-spouse-with-children count was set to zero.

2. An additional small number of cases could also be readily determined by looking at the spouse-spouse count. If one pair member had a spouse-spouse with children count that equaled or exceeded the final spouse-spouse count, but the other pair member

had a spouse-spouse with children count which was smaller than the final spouse-spouse count, then the final spouse-spouse with children count was set to the pair member's count that was consistent with the final spouse-spouse count.

For the remainder of general conditions, it had been established that at least one couple resided in the household:

3. For cases that were not already determined by looking at the previous two conditions, the counts for the two pair members (if there were two pair members) were equal in the vast majority of cases. The final count could be set to each pair member's count under the following conditions:

    - There was no potential for 2 or more couples in the household that were not already obviously identified, whereby one of the pair members had grandchildren and there were respondents 12 or over who were not children, grandchildren, siblings, children, parents, spouses, or partners, but no children-in-law were identified, and the number of household members was at least 4.
    - Two family units did not live in the household that were identified earlier as two family units.
    - Both pair members had valid rosters.
    - Either:
        - The counts were nonzero, and equal to the final spouse-spouse count, or
        - There were no bad relationship codes for roster members under 18, and one of the following held for at least one pair member:
            - The pair member's roster had no bad relationship codes for roster members 15 or over, or
            - The pair member was over 18, and had neither children nor siblings under 18 (covers zero counts, since no bad codes under 18), or
            - The pair member was under 18, did not have parents, but there was one bad relationship code among roster members over 18 in that pair member's roster (covers zero counts, since only one bad relationship code could potentially be a single parent, but not a pair of parents making a couple).

4. The pair members might both have zero counts, but the above conditions did not apply. The final count could still be zero if the age counts for both pair members and the screener indicated nobody lived in the household who was under 18, and there were no bad roster ages.

5. The counts for both pair members might still agree with nonzero counts, even though none of the previous conditions applied. The final count could still be set to one of the pair member's counts if the pair relationship was imputed to be a spouse-spouse pair with children, and there was no potential for 2 or more couples in the household that were not already obviously identified, whereby one of the pair members had

grandchildren and there were respondents 12 or over who were not children, grandchildren, siblings, children, parents, spouses, or partners, but no children-in-law were identified, and the number of household members was at least 4.

6. If one pair member did not have a valid roster but the other member did, the final count was set to the other pair member's count under the following conditions:

   - There was no potential for 2 or more couples in the household that were not already obviously identified, whereby one of the pair members had grandchildren and there were respondents 12 or over who were not children, grandchildren, siblings, children, parents, spouses, or partners, but no children-in-law were identified, and the number of household members was at least 4.
   - Either:
     - The count for the pair member with the valid roster was nonzero, and equal to the final spouse-spouse count, or
     - There were no bad relationship codes for roster members under 18, and one of the following held for the pair member with the valid roster:
       - The pair member's roster had no bad relationship codes for roster members 15 or over, or
       - The pair member was over 18, and had neither children nor siblings under 18 (covers zero counts, since no bad codes under 18), or
       - The pair member was under 18, did not have parents, but there was one bad relationship code among roster members over 18 in that pair member's roster (covers zero counts, since only one bad relationship code could potentially be a single parent, but not a pair of parents making a couple).

7. The pair member with the valid roster might have a zero count, but the above conditions did not apply. The final count could still be zero if the age counts for both the pair member with the valid roster and the screener indicated nobody lived in the household who was under 18, and there were no bad roster ages.

8. If the spouse-spouse-with-children counts disagreed in the same manner as the spouse-spouse counts disagreed, then the choice is obvious: use the count that corresponded to the correct spouse-spouse count. Details follow:

   - If the spouse-spouse-with-children counts were equal to the spouse-spouse counts for both pair members, even though they were unequal to each other, then the final spouse-spouse-with-children count was set to the final spouse-spouse count.
   - If the spouse-spouse counts exceeded the spouse-spouse-with-children counts by one for each pair member, even though they were unequal to each other, then the final spouse-spouse-with-children was set to one less than the final spouse-spouse count.

9. If two different family units had already been identified in the household, then two different parent-sets (one often a single parent) were being referenced. The final count was set to the sum of the two counts (where one of the counts was often zero).

10. Based on earlier conditions, we have already excluded households without couples. We have also excluded households with a possibility of two or more couples. If the pair relationship was parent-child, and at least one count was nonzero, then the identified couple must correspond to the parent-child relationship. The maximum of the counts was selected under the following conditions:

    - The sum of counts from the two pair members was 1.
    - There was no potential for 2 or more couples in the household that were not already obviously identified, whereby one of the pair members had grandchildren and there were respondents 12 or over who were not children, grandchildren, siblings, children, parents, spouses, or partners, but no children-in-law were identified, and the number of household members was at least 4.
    - Either:
        - The relationship was parent-child where the child was between 12 and 17, or
        - The relationship was parent-child where the child was between 18 and 20, and the child had siblings under 18.

11. Two couples have been identified in the household, where the household is multigenerational (one member of the younger couple is in a parent-child relationship with the older couple). If a sibling to the pair member in the younger couple was selected, or if a member of the younger couple was selected who "married into" the family, then he or she was not be able to identify the nephews, nieces, brothers-in-law, or sisters-in-law that could point to an appropriate accounting of all the couples with children, because of the relationship codes that were available. The maximum of the two counts was selected under the following conditions:

    - There was no potential for 2 or more couples in the household that were not already obviously identified, whereby one of the pair members had grandchildren and there were respondents 12 or over who were not children, grandchildren, siblings, children, parents, spouses, or partners, but no children-in-law were identified, and the number of household members was at least 4.
    - There were two couples in the household, as identified by the final spouse-spouse count.
    - The difference between the pair members' counts was 1.
    - Either:
        - The pair member with the smaller count had a spouse or partner, and the pair member with the larger count had parents in the household, or

- The pair member with the smaller count had parents-in-law or children-in-law in the household.

12. If a couple is a marriage/partnership that occurred after an earlier marriage, the partner may not consider the partner's children as his or her children, but the child, who was also selected, considered the spouse/partner as a parent. Even though the pair relationship is not parent-child, these cases are still counted as spouse-spouse with children since they are the children of one spouse/partner. The maximum count is selected under the following conditions:

    - There was no potential for 2 or more couples in the household that were not already obviously identified, whereby one of the pair members had grandchildren and there were respondents 12 or over who were not children, grandchildren, siblings, children, parents, spouses, or partners, but no children-in-law were identified, and the number of household members was at least 4.
    - One count was zero and the other count was one.
    - The pair member with the zero count had a spouse or partner.
    - The pair member with the nonzero count had parents.

13. The counts may have been unequal because children under 18 left, entered, or otherwise materialized or disappeared in the household after screening and between the time of the interviews. In general, the count was selected that corresponded to the pair member with a household composition closest to the screener household composition. If one pair member did not have children in the household, and the other pair member did, the following conditions were required for the count corresponding to the pair member with a household composition closest to the screener:

    - There was no potential for 2 or more couples in the household that were not already obviously identified, whereby one of the pair members had grandchildren and there were respondents 12 or over who were not children, grandchildren, siblings, children, parents, spouses, or partners, but no children-in-law were identified, and the number of household members was at least 4.
    - One pair member had a nonzero count of children under 18, and the other pair member had a zero count of children under 18.
    - Either:
        - The screener composition indicated children under 18 were in the household, whereupon the nonzero count was selected, or
        - The screener composition indicated no children under 18 were in the household, whereupon the zero count was selected.

14. The counts may have been unequal with a zero count and a count of one because a pair member with a zero count was not part of the immediate family unit. The nonzero count was used under the following conditions:

- There was no potential for 2 or more couples in the household that were not already obviously identified, whereby one of the pair members had grandchildren and there were respondents 12 or over who were not children, grandchildren, siblings, children, parents, spouses, or partners, but no children-in-law were identified, and the number of household members was at least 4.
- The pair relationship was not a parent-child, sibling-sibling, spouse-spouse, or grandparent-grandchild relationship.
- Both pair members had relationship codes that were not parent, child, sibling, spouse, partner, grandparent, or grandchild codes, among roster members who were 12 or over.

The following additional requirement was included, which overly restricted the cases that could be included within this general condition:

- The pair member with a nonzero count was under 21, and had 2 parents.

15. The counts may have been unequal because of bad relationship codes among roster members under 18. The following rules were used to determine if the count associated with the pair member who did not have bad relationship codes:

    - The number of roster members under 18 was the same between both pair members.
    - The side with the smaller count had one bad relationship code for roster members under 18.
    - There was no potential for 2 or more couples in the household that were not already obviously identified, whereby one of the pair members had grandchildren and there were respondents 12 or over who were not children, grandchildren, siblings, children, parents, spouses, or partners, but no children-in-law were identified, and the number of household members was at least 4.

16. If, after considering all of the general conditions given above, the count was left to imputation, it was still possible that the lower and upper bounds were equal. In this instance, the final count was set to one of the bounds.